Optimized kinematic models for the solar magnetic cycle

WHITBREAD, TIMOTHY,JAMES,OLIVER

Optimized kinematic models for the solar magnetic cycle

Tim Whitbread

A Thesis presented for the degree of
Doctor of Philosophy

Department of Mathematical Sciences
Durham University
United Kingdom

July 2019

Optimized kinematic models for the solar magnetic cycle

Tim Whitbread

Submitted for the degree of Doctor of Philosophy

July 2019

Abstract: Modelling the solar magnetic cycle requires the prescription of several poorly-constrained parameters. Accurate simulations are desirable because the state of the magnetic field at cycle minimum can be used to make predictions about the following cycle.

Small-scale parameter surveys have previously been performed in this area, but usually not with global coverage of the parameter space. In this thesis, a genetic algorithm is used to systematically search for optimal parameters for 1D and 2D surface flux transport models, with a view to applying the same technique to a kinematic 3D dynamo model. The method successfully obtains good matches with observations when applied to surface flux transport models. However, for more complex models a more efficient method might be needed. Such a method is Bayesian emulation and history matching, so we apply this method to the surface flux transport models and successfully recreate our results.

The contributions of individual active regions to the Sun's axial dipole moment are assessed by simulating the evolution of each region separately from the others. It transpires that a small number of active regions can have a significant effect on the end-of-solar-cycle dipole moment and hence the subsequent cycle. However, the cumulative effect of less important regions should not be ignored. Emergence latitude is the primary property of an active region in determining its contribution

to the axial dipole moment.

Finally, a discrepancy between the surface evolution in the surface flux transport model and dynamo model is investigated using a simple 2D diffusion model. The difference is due to radial diffusion which is not present in the surface-only model. An improved, yet suboptimal, match is obtained when either the diffusivity in the convection zone is increased, or the field lines of active regions are manually disconnected from the underlying toroidal field. Increasing diffusivity is a means of disconnecting regions from the toroidal field whilst conserving flux. However, it does not yet appear possible to maintain the dynamo with such a strong diffusivity, although a more thorough parameter optimization could solve this problem.

Declaration

The work in this thesis is based on research carried out in the Department of Mathematical Sciences at Durham University. No part of this thesis has been submitted elsewhere for any degree or qualification.

Acknowledgements

I would firstly like to thank my supervisor, Anthony Yeates, for his unwavering support and guidance over the past four years. I would not have made it this far without his patience in debugging code, improving readability of our publications, and providing rapid solutions to all of my problems, as well supplying everything I needed to carry out my research and complete my degree programme. My thanks extend to the other members of the MHD group at Durham University, past and present, for putting up with my dynamo talks and for unparalleled discussions about the food choices at the Palatine Centre.

I would also like to thank Andrés Muñoz-Jaramillo, my unofficial, remote secondary supervisor, for our enjoyable quasi-periodic Skype meetings and for his insightful and innovative suggestions on how to improve various aspects of my work. I am eternally grateful to Ian Vernon for his statistical advice, and for essentially supervising the work done in Chapter 3. Thank you also to Djoko Wirosoetisno, who stepped in as my secondary supervisor, despite having completely different research interests!

Additionally, I thank Gordon Petrie for providing the magnetogram data shown in Figure 1.6 which were acquired by SOLIS instruments operated by NISP/NSO/AURA/NSFI, and for his contributions to the first paper. I am grateful to other providers of data and numerical code, including David Hathaway, Paul Charbonneau, Barry Knapp and Travis Metcalfe. I apologise to anyone else I pestered with emails, particularly during the early stages of my studies which were governed by naïvety and youthful over-enthusiasm.

I would like to thank Durham University Department of Mathematical Sciences for

funding my PhD studentship, as well as the person who dropped out at the last minute to unknowingly ensure funding was available for me. My time at Durham would not have been the same without the vast array of characters I got to know at the University, especially those whom I shared an office with in my first year. I will miss our conversations about football, cricket, cultural differences, food, or anything else completely non-mathematical, and I am sure you will all go on to do great things in the future. I hope we can meet again some day.

Penultimately, I would like to thank those who gave me moral support throughout the past four years, including housemates, brothers and sisters in Christ, groundhoppers, and quiz team members. I also would like to thank members of my family for encouraging and being there for me, especially my parents who have supported me emotionally and financially regardless of my life choices, even though they will never truly understand the point in studying Mathematics. I am indebted to my wonderful wife Abby, who ambitiously suggested that I apply for a post-doc job in Durham, which eventually resulted in the PhD offer from Anthony. In all seriousness, being your husband has been a joy, even if planning a wedding whilst doing a PhD was a little stressful. Thank you for making me a better and happier person day by day.

Finally, my unending thanks go to my heavenly Father, who sent his Son to die in my place, and who created such an awesome universe for us to model numerically.

For the Lord gives wisdom; from his mouth comes knowledge and understanding.

— Proverbs 2:6

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Sun is a large mass of incandescent gas which sits at the centre of the Solar System. Even before heliocentrism became generally accepted as the astronomical model of our stellar system, humans were aware of the Sun's light- and life-giving energy and took time to study it in detail, recording its various temporal behaviours, even worshipping it as a deity. It is no wonder, then, that research of the Sun remains a pivotal area in astrophysics.

The Sun has a radius of $R_\odot \approx 696\,000\,\text{km}$ (Emilio *et al.*, 2012; Mamajek *et al.*, 2015), and much like the Earth its interior is made up of different layers (Figure 1.1). In the centre is the core, which is responsible for the production of nuclear energy via the fusion of hydrogen to form helium. In the radiative zone, the energy generated in the core is transported outwards from $\sim 0.25\,R_\odot$ by radiation, before reaching the convection zone at $\sim 0.7\,R_\odot$. Here the temperature gradient is too sharp to maintain static equilibrium, so an instability is induced and convective motions take control, transporting the energy to the surface (Tobias, 2002). In Figure 1.2 we see granulation on the solar surface caused by convection, with granule diameter being of the order 1000 km. The lighter areas represent convective upflows, whilst the dark, thin lines represent convective downflows. Larger granular patterns known as mesogranulation ($\sim 5000\,\text{km}$) and supergranulation ($\sim 32\,000\,\text{km}$) have been observed (Rast, 2003; Rieutord & Rincon, 2010), the latter of which is the most

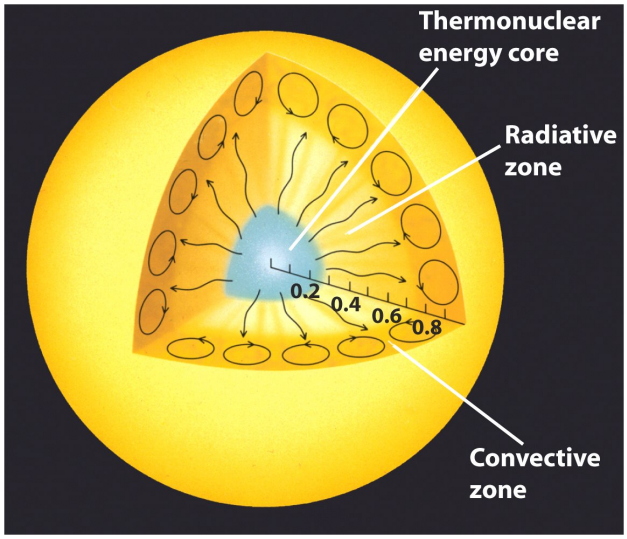

Figure 1.1: The solar interior. Energy is transported radially outwards due to radiation before convective motions take over at approximately $0.7\,R_\odot$. Taken from http://tuttidentro.files.wordpress.com.

important for the topics studied in this thesis, and is shown in Figure 1.3. This image was captured by Doppler imaging (Hart, 1954); red regions are where plasma is moving away from the viewer and blue areas are where material is moving towards the viewer. This is a line-of-sight image, indicating that motion is predominantly horizontal since the signal is weaker at the centre.

The solar atmosphere also consists of three 'layers'. The familiar visible surface is called the photosphere, where we observe features such as sunspots and convective granulation. Above this is the chromosphere where, using extreme ultraviolet and Hα images, we can observe bright plage regions, filamentary structures and prominences. The outermost atmosphere is called the corona, where temperatures rise to above 10^6 K, much hotter than the 20 000 K of the chromosphere and 6000 K of the photosphere. This phenomenon is surprising, and the so-called 'coronal heating problem' remains a major topic of debate in solar physics and magnetohydrodynamics (MHD). We know that the corona is heated ultimately by the conversion of magnetic energy (De Moortel & Browning, 2015), but the precise mechanism is still unknown. The most prominent theories are heating either by waves or by (nano-)flares.

In any case, observations of the atmosphere have aided us in understanding the

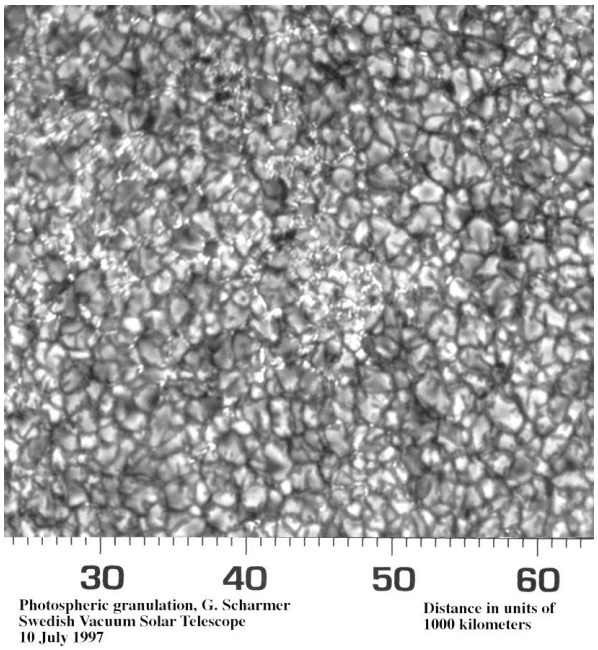

Figure 1.2: Granulation pattern on the surface of the Sun caused by small-scale convection of plasma. Lighter areas represent convective upflows and dark areas represent convective downflows. Obtained using the Swedish Vacuum Solar Telescope. Taken from http://solarscience.msfc.nasa.gov.

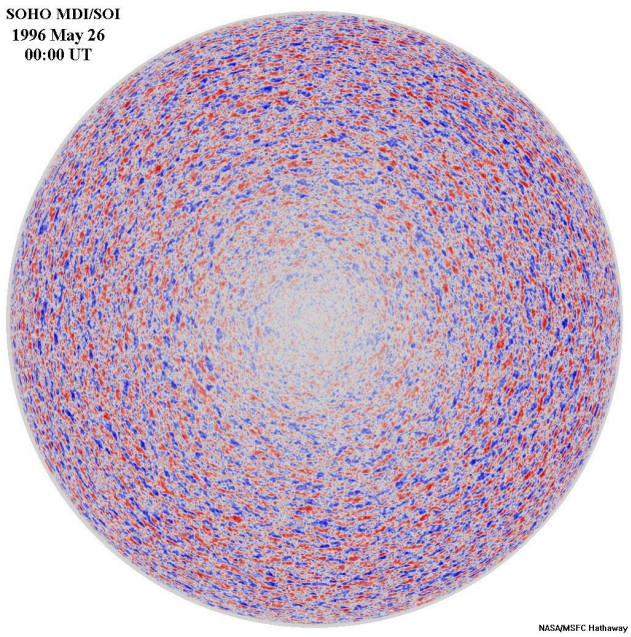

Figure 1.3: Supergranulation on the surface of the Sun indicating large-scale convection. Obtained using Doppler imaging of the photospheric velocity from SOHO MDI/SOI. Taken from http://solarscience.msfc.nasa.gov.

process by which flares and coronal mass ejections (CMEs) occur (e.g. van Driel-Gesztelyi & Culhane, 2009; Chen, 2011; Benz, 2017). While the specific mechanism for these eruptions is still an active area of research, it is understood that the Sun's magnetic field is responsible. Flares and CMEs release emissions across the electromagnetic spectrum into space through the solar wind. If these solar energetic particles (SEPs) reach the Earth, they interact with our magnetic field. This interaction can have positive impacts such as the famous aurora, but can also induce strong electric currents and disturbances in the ionosphere, affecting power grids and communication/signal transmission respectively. The effects of SEPs and magnetic disturbances on the Earth are collectively known as 'space weather'.

1.1 The solar cycle

With the introduction of space weather onto the UK National Risk Register[a], the importance of understanding the impact of magnetic activity in the Solar System has never been higher. In particular, the behaviour of the Sun's magnetic field can have adverse effects on technology and business, as observed during the 'Carrington event', a large solar storm in 1859 which, while beautiful and spectacular through the manifestation of aurorae, wiped out telegraph networks across the globe (Carrington, 1859; Hodgson, 1859; Green *et al.*, 2006). It is estimated that the disruption caused by an event of similar magnitude today could cost the US alone $ 1–2 trillion, with the global economy facing a loss of up to $ 3.4 trillion (Schulte in den Bäumen *et al.*, 2014; Schrijver, 2015; Schrijver *et al.*, 2015). Moreover, the ejection of highly energized particles would be harmful for any space-based astronauts outside the protection of the Earth's magnetic field. Additional financial losses could arise with the destruction of satellites and other spacecraft. After having researched a variety of historical records, Schrijver & Beer (2014) concluded that even more energetic

[a]https://www.gov.uk/government/publications/national-risk-register-for-civil-emergencies-2015-edition

events are possible.

To predict when a catastrophic event could reoccur requires robust models for the evolution of the Sun's magnetic field. These in turn require understanding of the Sun's key physical features, as well as the regular recording of observational data. The most accessible data come from sunspot records which have been collected for hundreds of years.

Sunspots are cooler, darker patches on the solar surface which arise due to the presence of strong magnetic fields ($\sim 3000\,\mathrm{G}$), a feature discovered by George Ellery Hale (Hale, 1908). While sunspots themselves have been recorded for centuries, with earliest known recordings as far back as around 300 BC (Vaquero, 2007), Hale noticed that the Zeeman effect, that is, the splitting of spectral lines due to a magnetic field, occurs in the umbra of a sunspot. The magnetic field suppresses convection, resulting in an area of relatively low temperature ($\sim 3800\,\mathrm{K}$, compared to 6000 K, the temperature of the surrounding photosphere) and hence a darker region.

Sunspots follow a cyclic pattern with a period of approximately eleven years, first documented by Heinrich Schwabe (Schwabe, 1844). At the start of a cycle, sunspots appear at latitudes of $\pm 30°$ in each hemisphere. As the cycle progresses, regions of sunspot emergence (and hence emerging magnetic fields) migrate equatorward, and by the end of the cycle sunspot emergence is generally restricted to around $\pm 5°$. This pattern was discovered by Richard Carrington (Carrington, 1863), but is known as 'Spörer's Law' after the astronomer Gustav Spörer (Spörer, 1880), who refined the work of Carrington. This cycle can be seen in the so-called 'butterfly diagram' (Figure 1.4) which shows latitudes of sunspot emergence over time.

The sunspot number is defined as the sum of the number of individual identifiable sunspots and ten times the number of sunspot groups. This provides a reliable estimate to the number of sunspots because there are on average ten spots within a group. The sunspot number varies significantly between cycles, as shown in Figure 1.5 (Hathaway, 2010). What is striking in this figure is the decreased levels of sunspot

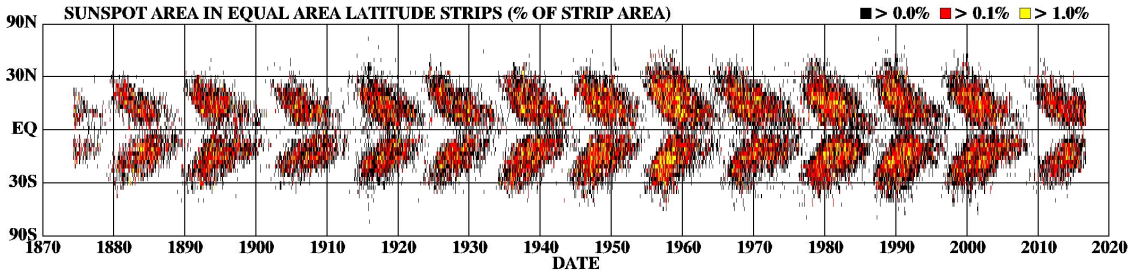

Figure 1.4: Butterfly diagram showing the latitude and emergence time of sunspots. Colour represents sunspot area in each latitudinal strip. Taken from http://solarscience.msfc.nasa.gov.

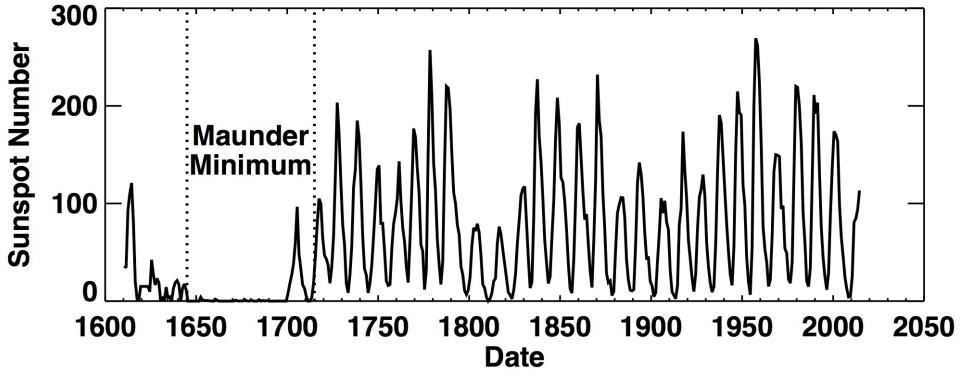

Figure 1.5: Sunspot number over time. The Maunder Minimum was a period of decreased magnetic activity. Taken from http://solarscience.msfc.nasa.gov.

incidence between 1634 and 1715. This period is known as the 'Maunder Minimum' (Eddy, 1976), and was a genuine phenomenon where solar magnetic activity was at a low. Abundances of ^{10}Be and ^{14}C, which can be found in polar icecaps and tree rings respectively, are anti-correlated with magnetic activity since they are produced by cosmic rays entering the Earth's atmosphere, which are in turn modulated by the solar wind. Analysis of these isotopes show grand minima similar to the Maunder Minimum have occurred multiple times over thousands of years, with a mean period of approximately 200 years (Beer, 2000). Magnetic activity continued during these minima, but at a reduced level (Beer *et al.*, 1998).

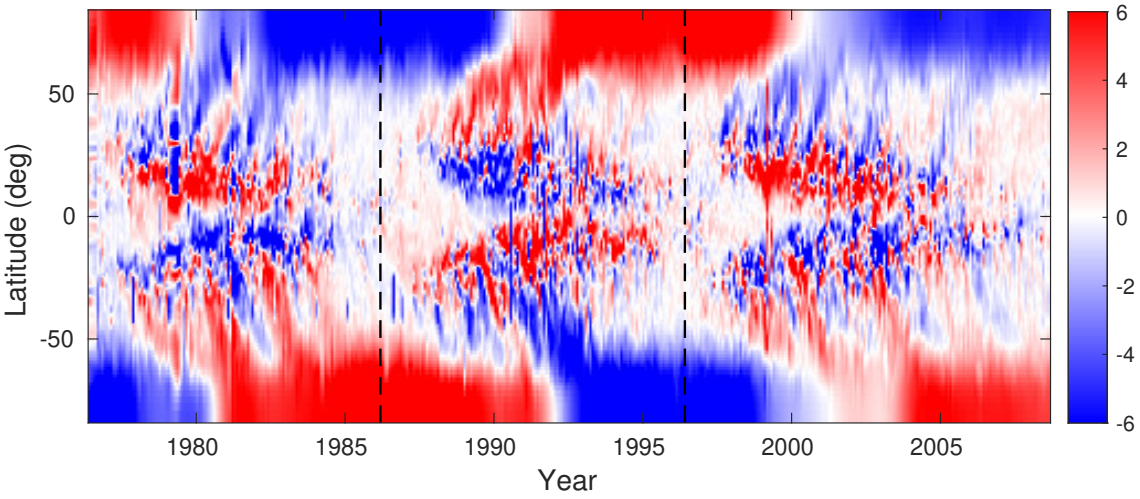

Figure 1.6: Magnetic butterfly diagram from Kitt Peak magnetograms highlighting field reversal between solar cycles 21–23. Vertical dashed lines represent start/end points of cycles as used in this thesis. Red is positive polarity, blue is negative polarity, and units are in G.

1.2 The solar dynamo

Following Hale's findings, the father and son team of Harold and Horace Babcock used their newly-developed magnetograph to show that magnetic fields are ubiquitous on the Sun, with a predominantly dipolar large-scale field (Babcock & Babcock, 1955). Thanks to the aforementioned discoveries we can now obtain full-disk images of the Sun's magnetic field at the surface and track its evolution, investigate large- and small-scale structures, or piece the images together to form magnetic butterfly diagrams, such as the one shown in Figure 1.6.

Figure 1.6 was obtained using full-disk images from the US National Solar Observatory, Kitt Peak, which underwent a polar field correction procedure described by Petrie (2012). Red and blue areas represent positive and negative polarities respectively, which reverse at the end of each sunspot cycle. It is clear that the 11-year sunspot cycle is actually due to a 22-year magnetic cycle (Hale, 1924). It is now generally believed that this magnetic cycle is maintained by a magnetohydrodynamic dynamo operating in the convection zone (Larmor, 1919). A dynamo is the process which describes the regeneration of the magnetic field, usually through some

interaction between the field and the background flows. The magnetic field can be decomposed into its toroidal (azimuthal) and poloidal (meridional) components, and we tend to view the dynamo process as the cyclical generation of one component from the other (Tobias, 2002).

Unfortunately we cannot observe or measure the magnetic field inside the Sun. Nevertheless we have some indication of the dynamic processes occurring beneath the surface thanks to helioseismology, the study of acoustic wave oscillations in the Sun. In particular, the differential rotation profile of a large portion the solar interior has been mapped out (Howe, 2009), as shown in Figure 1.7. The angular velocity in this figure was obtained by applying a 1.5D regularized least squares inversion technique to frequency splittings from observed oscillations in the solar interior, measurements of which were made via the *Solar and Heliospheric Observatory (SOHO)* Michelson Doppler Imager (MDI; Scherrer *et al.*, 1995). This method is described in thorough detail by Schou *et al.* (1998), but briefly, the angular velocity is expanded as:

$$\Omega\left(r,\theta\right)=\sum_{s=0}^{s_{\max}}\Omega_s(r)\,\psi_{2s}^{(1)}(x), \tag{1.2.1}$$

where $x\equiv\cos\theta$ and the functions $\psi_{2s}^{(1)}(x)$ are chosen to be:

$$\psi_{2s}^{(1)}(x)=\frac{dP_{2s+1}}{dx}, \tag{1.2.2}$$

where $P_k\left(x\right)$ are Legendre polynomials (see Chapter 2).

It is clear that the surface rotation rate is faster at the equator than at higher latitudes, and this trend continues throughout the convection zone. The radiative zone, however, rotates more as a solid body, and there is a transition region between which has the strongest radial shear, known as the tachocline (Spiegel & Zahn, 1992). This thin, stably-stratified layer plays an important role in the dynamo process, as described below. While the differential rotation is approximately stable over time, small temporal variations have been observed (Howard, 1983). The origins and dynamics of these torsional oscillations are still under investigation but they could also play an important role in dynamo theory (Rempel, 2007; Guerrero *et al.*, 2016).

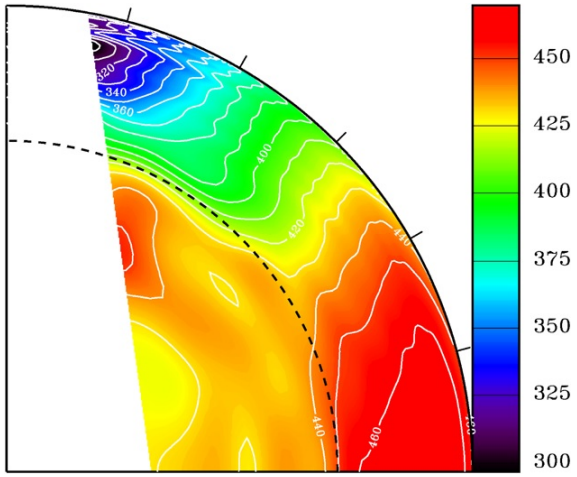

Figure 1.7: Meridional cut of the solar interior showing rotation rate at different radii and latitudes. Contours are labelled in nHz. The dashed line represents the base of the convection zone, and tick marks highlight latitudes of 15°, 30°, 45°, 60° and 75°. Taken from Schou *et al.* (1998).

Differential rotation contributes by stretching poloidal field azimuthally, so that sections of field lines near the equator will be stretched further than those at slowly-rotating higher latitudes. This produces toroidal field and provides the mechanism for one stage of the dynamo process, known as the ω-effect. Converting toroidal field back to poloidal field is not so straightforward. Parker (1955*a*) suggested that the toroidal field is stretched radially by convective upflows then twisted by the Coriolis effect to form small-scale poloidal loops. The net effect of these loops is a large-scale poloidal field, as required. This is called the α-effect.

The simple solar dynamo recipe is then as follows: turbulent pumping (convective downflow) sends poloidal flux to the tachocline at the base of the convection zone (Tobias *et al.*, 1998), where the effect of differential rotation is very strong (some models instead rely on meridional flow and/or turbulent diffusion to transport the poloidal flux down to the tachocline). Regions of intense magnetic flux are less dense than their surroundings and hence become buoyant (Parker, 1955*b*). If they were stored in the convection zone, they would rise to the surface on a shorter timescale than the 11 years we observe in butterfly diagrams. Furthermore, the diffusion rate in the turbulent convection zone may be too strong for field to be generated

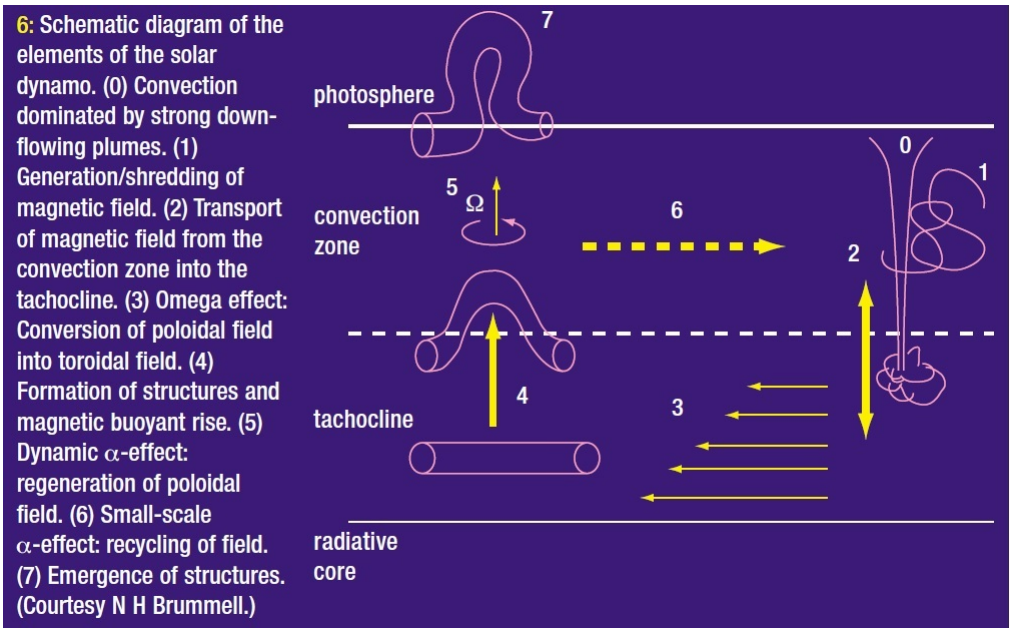

Figure 1.8: A sketch outlining the dynamo process. Taken from Bushby & Mason (2004).

or maintained. It seems sensible therefore to place the dynamo at the base of the convection zone where the turbulent diffusion effect is smaller. The poloidal field is then stretched into toroidal field via the ω-effect, and this is transported back to the convection zone by magnetic buoyancy and diffusion, where it is in turn deformed by the α-effect to produce poloidal field. The poloidal field is carried down to the tachocline again and the cycle continues. Only the strongest magnetic field will be carried to the surface to form sunspots. This process is sketched in Figure 1.8.

There are numerous varieties of dynamo models in the literature, each with their own strengths and limitations, but in this thesis we will focus on Babcock-Leighton (B-L) dynamos, which are a type of flux transport dynamo (FTD). The appeal of B-L dynamos is that they rely on the decay of sunspots ('active regions') at the surface, as well as interior and surface flows, all of which can be observed to some degree. Furthermore, they have been found to reproduce features of the solar cycle discussed in Section 1.1 (Dikpati *et al.*, 2004; Mackay & Yeates, 2012). However, the reliance on sunspots can also be a problem, particularly if a period of grand minimum is reached when there is a dearth of sunspots and the dynamo can no longer operate, although Karak & Miesch (2018) demonstrated that with a strong

enough turbulent pumping mechanism a 3D FTD model was able to escape from grand minimum periods.

A distinction between B-L dynamos and some other types of dynamo is that they do not require an explicit turbulent α-effect. This mechanism is usually expressed mathematically via mean-field electrodynamics, that is, the separation of large (mean) and small (fluctuating) scales which has been the driving force of dynamo theory for the best part of half a century (Moffatt, 1978; Parker, 1979; Krause & Rädler, 1980), but open problems remain regarding the mechanism for transporting appropriate amounts of flux to and from the tachocline, or avoiding the quenching of the α-effect in the presence of strong magnetic fields. For a recent comprehensive review of other types of dynamo and dynamo theory in general, see Charbonneau (2014).

1.2.1 The Babcock-Leighton mechanism

Sunspots generally appear in pairs corresponding to bipolar magnetic regions (BMRs), which appear on the solar surface due to the rise of buoyant flux ropes which break through the photosphere (Fan, 2009; Cheung & Isobe, 2014). BMRs tend to emerge tilted at an angle with respect to the east-west line (using the line connecting the centres of the opposing polarities), with the leading polarity emerging closer to the equator. This is thought to be due to the Coriolis effect in the convection zone and is more pronounced at higher latitudes, according to Joy's law (Howard, 1991). The leading polarities are antisymmetric about the equator for each cycle, and at the end of the cycle the polarity reverses, resulting in the 22-year magnetic cycle.

Robert Leighton conjectured that magnetic flux could be transported on a random walk process over the supergranular convective cells on the solar surface (Leighton, 1964), which are shown in Figure 1.3. In the continuum limit this discrete random walk process becomes supergranular diffusion. The leading polarity flux diffuses across the equator and cancels out with the corresponding opposite flux from the

other hemisphere. This leaves net trailing polarity flux to be transported poleward by diffusion, as indicated by the off-vertical 'surges' shown at mid-latitudes in Figure 1.6 (Leighton, 1965). The trailing flux neutralizes the remaining opposite-flux polar field from the previous cycle and stores itself at the pole instead. Polar field reversal typically occurs around sunspot maximum.

However, the diffusion process alone is too slow to account for the timing of reversals in the polar field (Sheeley, 2005) and most models now include a meridional 'conveyor belt' flow which aids poleward flux transport at the surface, and gives rise to the poleward surges observed in magnetograms. This was the original idea of Horace Babcock (Babcock, 1961), who in turn was missing the diffusion aspect of Leighton's model, which is important for cross-equatorial cancellation. The return meridional flow, thought to be at the base of convection zone, is a possible explanation for the equatorward migration of sunspot emergence and helps induce polarity reversal (Wang *et al.*, 1991; Choudhuri *et al.*, 1995). Meridional circulation is a relatively slow transport mechanism, with surface speeds of $\sim$10–20 $\mathrm{m\,s^{-1}}$ observed via helioseismology (e.g. Braun & Fan, 1998; Zhao & Kosovichev, 2004; Jackiewicz *et al.*, 2015). However, helioseismic recordings are near the limit of credibility (Komm *et al.*, 2013) and so the flow profile is not well constrained. Flow speeds within the same range have also been found by e.g. Komm *et al.* (1993) via the tracking of small magnetic features, and by Topka *et al.* (1982) via the comparison of polar zone filament distribution and polar magnetic field evolution. Furthermore, many simulations have produced more accurate results when a meridional flow was included than when only diffusion was considered (DeVore *et al.*, 1985; Wang *et al.*, 1989*b*). It is now generally accepted that such a poleward flow exists – the big question is whether the flow is an important component in the dynamo. The combination of diffusion and meridional circulation form what is known as the Babcock-Leighton mechanism, the cyclic process in B-L dynamo models in which the decay of sunspots leads to the production of poloidal field. A cartoon of the B-L mechanism is shown in Figure 1.9.

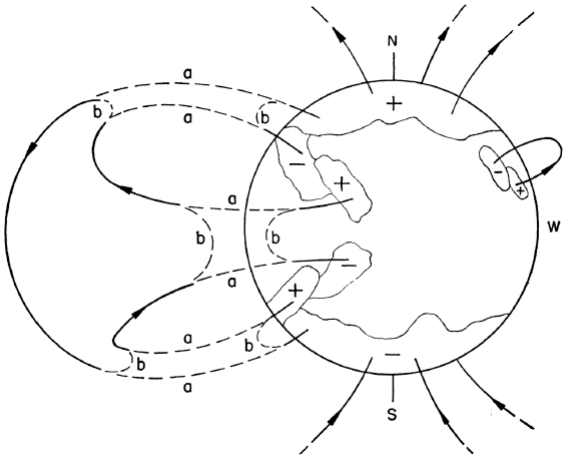

Figure 1.9: Sketch of the B-L process: the production of a large-scale poloidal field from the decay of active regions. Taken from Babcock (1961).

In some cases, it is sufficient to only consider the surface/photospheric component ($r = R_\odot$) of such models. Tracking and modelling magnetic regions on the solar surface, i.e. surface flux transport (SFT; Leighton, 1964), has been highlighted as a key method for predicting and understanding solar cycle variability (e.g. Upton & Hathaway, 2014*b*; Hathaway & Upton, 2016), without having to make assumptions about the solar interior where more limited observations are available to constrain dynamo models. This is because the strength of the polar field at the end of the cycle is found to be a good indicator of the strength of the following cycle in terms of the maximum sunspot number (e.g. Schatten *et al.*, 1978; Muñoz-Jaramillo *et al.*, 2013). Thus using some form of source term to represent the BMRs, we can easily simulate the subsequent evolution of magnetic regions due to advective and diffusive transport on the surface alone to derive an estimate for the end-of-cycle polar field strength, and hence the amplitude of the following cycle. In most SFT models BMR emergences at high latitudes are not taken into account, which can occur on occasion as found by Durrant *et al.* (2001, 2002), although they found that any discrepancies were small and that generally there was a good agreement between observations and models even when high-latitude regions were not considered. Some SFT models have included these features (Worden & Harvey, 2000; Schrijver, 2001).

One limitation of traditional SFT models (described mathematically in Section 1.3.2) is that there is no radial loss of magnetic flux. The consequence is that an excess of polar flux can build up, meaning reversal does not occur before the following cycle begins. Schrijver *et al.* (2002) combated this by including an extra exponential decay term in the model, which will be discussed in more detail throughout the thesis.

Cycle amplitude modulation can be achieved in SFT models by changing the meridional flow velocity. Counter-intuitively, if the poleward flow is slower it means that leading BMR polarities have longer to diffuse across the equator, leaving more trailing flux to reach the poles, resulting in a stronger polar field, and vice versa. Wang *et al.* (2002*a*) used this fact to maintain regular polar field reversals by varying the meridional flow speed from one cycle to the next. However, it is unclear whether a slow meridional flow is the cause of a strong cycle, or whether strong cycles quench the meridional flow.

Modulation can also be achieved by varying tilt angles of BMRs. Reducing the tilt inhibits cross-equatorial cancellation, meaning each active region will contribute less to the polar field. In extreme circumstances, abnormal orientation or deviations from Joy's Law might even be causes of grand minima (see Chapter 4). Cameron *et al.* (2010) were able to recover a more realistic polar field evolution by reducing the tilt angles in stronger cycles, with the physical reasoning being that stronger magnetic fields can resist deformation by the Coriolis effect (tilt- or α-quenching as mentioned above). However, we are not really at liberty to vary the tilt angles if we take more realistic representations of observed active regions in our models.

Of particular interest is the unusually weak polar field at the end of Cycle 23 (Muñoz-Jaramillo *et al.*, 2012), which in turn is believed to be responsible for the low amplitude of Cycle 24. This solar minimum period also lasted for longer than expected (Muñoz-Jaramillo *et al.*, 2015), and SFT models initially struggled to replicate the behaviour (Schrijver & Liu, 2008). Jiang *et al.* (2013) presented changes in either meridional flow speed or tilt angles as mechanisms for reproducing the weak polar field, but a definitive reason for the unusual behaviour is still sought after,

with some suggesting that the Sun is undergoing a transitional period in dynamo behaviour (Howe *et al.*, 2017) possibly leading to an extended grand minimum period. Incidentally, the strength of a cycle is anti-correlated with the duration of the cycle, according to the Waldmeier Effect (Wolf & Brunner, 1935). An explanation is offered by Cameron & Schüssler (2016): toroidal activity belts near the base of the convection zone cancel across the equator with the opposing polarity. Stronger cycles have wider belts and so cancellation occurs earlier, resulting in a shorter cycle since all declining phases are approximately the same.

1.2.2 Solar cycle prediction

As mentioned above, a major goal (or, perhaps, dream) of solar physics research is to be able predict the timing and amplitude of future solar cycles. Pesnell (2016) gave an overview of the predictions made for Cycle 24, around the time of its onset in late 2008 (see also McIntosh *et al.*, 1979, for predictions of Cycle 21; Brown, 1986, for Cycles 21 and 22; Layden *et al.*, 1991, for Cycle 22; and Li *et al.*, 2001, for Cycles 22 and 23). The prediction methods for Cycle 24 were separated into six categories: climatology, recent climatology (post Cycle 17), precursor, dynamo model, spectral, and neural networks. The precursor method, i.e. using the state of the Sun's polar magnetic field or geomagnetic activity (Vennerstrøm & Friis-Christensen, 1987) at solar minimum to form a prediction, was in general the most successful, provided the prediction was made close to cycle minimum. Using the polar field was particularly fruitful, and we will focus on this method specifically throughout this thesis.

Herein we advance this technique by optimizing parameters for SFT models, which can be used to simulate the end-of-cycle polar field. The output is calibrated against observed data and so will be as accurate as possible. We also assess the contribution of individual magnetic regions on the solar surface to the polar field, to see how important a role a small number of large regions could play in solar cycle variability and prediction.

The precursor method has already been utilised by some studies to predict the amplitude of Solar Cycle 25. Hathaway & Upton (2016) simulated the remainder of Cycle 24 using active region data from Cycle 14, exploiting the similarities between the two cycles. Multiple realizations were performed with changes to the meridional flow profile and Joy's Law tilt angle variations. They found an average axial dipole moment at the start of Cycle 25 of 1.36 ± 0.20 G compared to -1.61 G at the start of Cycle 24, 3.21 G at the start of Cycle 23, and -4.40 G at the start of Cycle 22. An updated prediction of 1.56 ± 0.05 G for the start of 2020 and 1.54 ± 0.04 G for the start of 2021 was made two years later (Upton & Hathaway, 2018). Note that the new prediction carries less uncertainty because it was made deeper into the declining phase of the cycle. As we will discuss in Chapter 4, this is because there are fewer regions emerging in the declining phase, making it less likely for a large region to emerge which could have a significant impact on the axial dipole moment. Nevertheless, the new prediction was within the error bars of the old one.

Cameron *et al.* (2016) performed a similar procedure, but with randomly generated active regions with properties drawn from empirical formulae. An end-of-cycle axial dipole moment of 2.5 ± 1.1 G was calculated, which is slightly stronger than the prediction of Hathaway & Upton (2016), but has a much wider bound of uncertainty. The forecast is still for a weaker cycle than Cycles 21–23. Other predictions made using the precursor method include 1.76 ± 0.68 G by Jiang & Cao (2018) and 2 G by Wang (2017).

Some recent studies have used alternative methods to form predictions of Cycle 25. For example, Hawkes & Berger (2018) used the correlation between magnetic helicity flux and cycle amplitude to predict that Cycle 25 will be of similar strength to or only slightly stronger than Cycle 24. However, it is not clear whether the current helicity flux cycle has yet reached its peak, and the correlation between helicity and following cycle strength is only based on a few data points. Consequently the uncertainty in the prediction is large. Gopalswamy *et al.* (2018) found that the polar microwave brightness in one cycle is correlated with the low latitude brightness with

a lag of about half a cycle, which is the microwave equivalent of the magnetic polar field precursor method. The subsequent conclusion was that Cycle 25 will be similar to Cycle 24.

Whilst these forecasts are generally consistent with each other, predicting that Cycle 25 will be another weak cycle, we will have to wait a few more years before their accuracy can be judged. If it is indeed a weak cycle, it may be preceded by a long, extended minimum period, so the true minimum may not be reached until 2021 (Upton & Hathaway, 2018). Additionally, Cycle 24 would then not be an isolated weak cycle, so could be the onset of an extended minimum period. Whether it is a short period like the Gleissberg minimum or deeper like the Dalton or even Maunder minimum remains to be seen.

1.3 Modelling the Sun's magnetic field

1.3.1 MHD equations

After discussing key features of the Sun and the dynamo process, we can now derive equations to model the Sun's magnetic activity. We will use the MHD equations (Priest, 1982). Although plasma is made up of different particle species, i.e. electrons and ions, MHD approximates it as a single fluid with magnetic field $\mathbf{B}$, electric field $\mathbf{E}$ and velocity field $\mathbf{u}$. We start with the pre-Maxwell equations:

$$\nabla \cdot \mathbf{B} = 0, \tag{1.3.1}$$

$$\nabla \cdot \mathbf{E} = \frac{\rho_c}{\varepsilon_0}, \tag{1.3.2}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}, \tag{1.3.3}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \tag{1.3.4}$$

Equation 1.3.1 is the solenoidal condition, that is, there are no magnetic monopoles or point sources of magnetism, and that magnetic flux is conserved. Conversely,

Equation 1.3.2 states that the net electric flux through a closed surface is proportional to the charge density ρ_c within that surface. The constant ε_0 is permittivity, the ability to store electrical energy in an electric field. However, because the length scales assumed in MHD are much longer than gyroradii of electrons and ions, and the Debye length, the distance over which a charge carrier's electrostatic effect still occurs, we can neglect Equation 1.3.2.

Equation 1.3.3 is Ampère's Law, which states that the magnetic field around a closed loop is related to the current density $\mathbf{j}$ within the loop. This applies in the case of static electric fields. Maxwell's addition was to recognise the extra effect contributed by time-varying electric fields. The equation then becomes:

$$\nabla \times \mathbf{B} = \mu_0 \left(\mathbf{j} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right), \tag{1.3.5}$$

where μ_0 is (constant) permeability, the ability to support the formation of a magnetic field. If we write $\mu_0 \, \varepsilon_0 = c^{-2}$, where c is the speed of light, then for non-relativistic systems this extra 'displacement current' term becomes very small and is neglected in MHD. Finally Equation 1.3.4 is Faraday's law of induction, which describes how a time-varying magnetic field can induce a voltage within a closed circuit.

We also use Ohm's Law in a moving medium,

$$\mathbf{j} = \sigma \left(\mathbf{E} + \mathbf{u} \times \mathbf{B} \right), \tag{1.3.6}$$

where σ is (constant) conductivity, the ability to conduct an electric current. Then,

$$\begin{aligned} \nabla \times \mathbf{j} &= \sigma \left(\nabla \times \mathbf{E} + \nabla \times (\mathbf{u} \times \mathbf{B}) \right) \\ &= \sigma \left(-\frac{\partial \mathbf{B}}{\partial t} + \nabla \times (\mathbf{u} \times \mathbf{B}) \right). \end{aligned} \tag{1.3.7}$$

Also we have (Equation 1.3.3):

$$\nabla \times (\nabla \times \mathbf{B}) = \mu_0 \left(\nabla \times \mathbf{j} \right), \tag{1.3.8}$$

and so we can use the identity $\nabla \times (\nabla \times \mathbf{G}) = \nabla (\nabla \cdot \mathbf{G}) - \nabla^2 \mathbf{G}$ to get

$$\nabla \times (\nabla \times \mathbf{B}) = -\nabla^2 \mathbf{B}, \tag{1.3.9}$$

since $\nabla \cdot \mathbf{B} = 0$.

Putting this together, we have:

$$-\frac{1}{\mu_0 \sigma} \nabla^2 \mathbf{B} = -\frac{\partial \mathbf{B}}{\partial t} + \nabla \times (\mathbf{u} \times \mathbf{B}), \tag{1.3.10}$$

or

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}, \tag{1.3.11}$$

where $\eta = \dfrac{1}{\mu_0 \sigma}$ is diffusivity. This is the MHD induction equation (or advection-diffusion equation), which describes how the magnetic field changes over time due to the effects of both advection with the fluid (first term) and diffusion (second term). Here we have assumed an isotropic plasma for simplicity, resulting in a constant conductivity and diffusivity. This may be a valid assumption at the photosphere where such properties do not vary significantly on average across the solar surface, but later we will prescribe some radial dependence to the diffusivity, and must use a more general form of the induction equation.

For completeness, the remaining MHD equations may be stated as follows:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{1.3.12}$$

$$\rho \left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \right) \mathbf{u} = -\nabla p + \mathbf{j} \times \mathbf{B} + \mathbf{F}, \tag{1.3.13}$$

$$\rho \left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \right) e + \rho (\gamma - 1) e \nabla \cdot \mathbf{u} = -\mathcal{L}. \tag{1.3.14}$$

Note that here ρ is fluid density and is distinct from ρ_c in Equation 1.3.2. The pressure is denoted by p and $\mathbf{F}$ is an additional force term including gravitational and viscous forces. The ratio of specific heats is given by γ and for the adiabatic case we take $\gamma = \frac{5}{3}$. Then the internal energy $e = \dfrac{p}{(\gamma - 1)\rho}$. The term on the right-hand side of Equation 1.3.14, $\mathcal{L}$, is the total energy loss function which includes contributions from ohmic heating, thermal conduction and radiation.

Equation 1.3.12 is the continuity equation, which states that mass within a closed system is conserved. Equation 1.3.13 is the momentum equation, describing the motion of the fluid. This is similar to the Navier-Stokes equation in hydrodynamics, but with an additional Lorentz force term $\mathbf{j} \times \mathbf{B}$, which represents the force imparted by the magnetic field on a charged particle. Finally, Equation 1.3.14 is the energy equation.

In the full problem, we would solve Equations 1.3.12–1.3.14 simultaneously with the induction equation. However, in this thesis we take a kinematic approach, where we prescribe the velocity $\mathbf{u}$ based on observations of the flows. It is therefore sufficient in this regime to solve only the induction equation to describe the evolution of the magnetic field in the convection zone and at the photosphere.

One final important quantity in MHD is the magnetic Reynolds number, a dimensionless number given by:

$$R_m = \frac{UL}{\eta}, \tag{1.3.15}$$

for some characteristic velocity scale U and length scale L. Typically in astrophysical scenarios $R_m \gg 1$ (e.g. $R_m \sim 10^6$ for the Sun), and so the effects of diffusion become negligible compared to advection, meaning we can remove dissipative terms in the MHD equations. This regime is called 'ideal MHD'. However, we will continue to include a turbulent diffusion term in the induction equation, in order to represent the diffusive effect of unresolved small-scale convective motions on the large-scale mean magnetic field. Such a diffusive term also aids the numerics.

1.3.2 Surface flux transport equation

Equation 1.3.11 can be written more generally as:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) - \nabla \times (\eta \nabla \times \mathbf{B}), \tag{1.3.16}$$

and we use spherical coordinates (r, θ, ϕ), representing the radial, co-latitudinal and longitudinal directions respectively. For modelling at the surface alone, we typically

assume the field to be exclusively radial (Martínez Pillet *et al.*, 1997). Then taking the radial component of Equation 1.3.16 and neglecting radial derivatives we have:

$$\begin{aligned}[\nabla \times (\mathbf{u} \times \mathbf{B})] \cdot \mathbf{e_r} &= [\nabla \times (u_\phi B_r \mathbf{e}_\theta - u_\theta B_r \mathbf{e}_\phi)] \cdot \mathbf{e_r} \\ &= -\frac{1}{R_\odot \sin\theta} \left(\frac{\partial}{\partial \theta} \left(u_\theta \, \sin\theta \, B_r \right) + \frac{\partial}{\partial \phi} \left(u_\phi B_r \right) \right), \end{aligned} \tag{1.3.17}$$

and

$$\begin{aligned}[\nabla \times (\eta \nabla \times \mathbf{B})] \cdot \mathbf{e_r} &= \left[\nabla \times \left(\frac{\eta}{R_\odot \sin\theta} \frac{\partial B_r}{\partial \phi} \mathbf{e}_\theta - \frac{\eta}{R_\odot} \frac{\partial B_r}{\partial \theta} \mathbf{e}_\phi \right) \right] \cdot \mathbf{e_r} \\ &= -\frac{\eta}{R_\odot^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \, \frac{\partial B_r}{\partial \theta} \right) - \frac{\eta}{R_\odot^2 \sin^2\theta} \frac{\partial^2 B_r}{\partial \phi^2}, \end{aligned} \tag{1.3.18}$$

since in this thesis we shall only consider $\eta \equiv \eta\,(r)$ or $\eta =$ constant. Factors of $1/\sin\theta$ arise as a result of taking the curl in spherical coordinates. This produces singularities at the poles which could cause numerical problems. However, no problems arise because we use a staggered grid where values of B_r are defined on the centres of the cell faces, and therefore offset from the pole. The boundary conditions at the poles on the grid ensure magnetic flux conservation.

Putting Equations 1.3.17 and 1.3.18 together, we have:

$$\begin{aligned} \frac{\partial B_r}{\partial t} &= \frac{\eta}{R_\odot^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \, \frac{\partial B_r}{\partial \theta} \right) + \frac{\eta}{R_\odot^2 \sin^2\theta} \frac{\partial^2 B_r}{\partial \phi^2} \\ &\quad - \frac{1}{R_\odot \sin\theta} \left(\frac{\partial}{\partial \theta} \left(u_\theta \, \sin\theta \, B_r \right) + \frac{\partial}{\partial \phi} \left(u_\phi B_r \right) \right), \end{aligned} \tag{1.3.19}$$

which is known as the surface flux transport equation. It was first presented by Leighton (1964) and has been in frequent use since (e.g. DeVore *et al.*, 1984; Wang *et al.*, 1989*a*; van Ballegooijen *et al.*, 1998; Schrijver & Title, 2001; Baumann *et al.*, 2004; Sheeley, 2005; Jiang *et al.*, 2010). Nevertheless, the model has changed over the years since Leighton's diffusion-based version: meridional flow is now generally included and the diffusion is typically much weaker, as we shall discuss in Chapter 2. Due to increased computational power, we can now simulate large numbers of magnetic regions at a time, in stark contrast to the early simulations of individual regions (Sheeley *et al.*, 1983). The SFT model has also been coupled to coronal field

models (Wang *et al.*, 1988), and used to simulate the magnetic field of other stars (Mackay *et al.*, 2004). In Chapter 2 we will add more details to Equation 1.3.19 involving velocity, diffusion, and sinks and sources.

1.4 Thesis outline

In this thesis we investigate models of solar magnetic field evolution, namely SFT and dynamo models, and attempt to calibrate their parameters against observed data. We also assess the contribution of individual source terms to the final polar magnetic field, and explore inherent differences between the two types of model.

In Chapter 2 we present a 1D model for surface flux transport and the genetic algorithm used to perform the optimization. We repeat the optimization process on a 2D SFT model with an automated region identification and assimilation process (Yeates *et al.*, 2015), and use this to search for variation in input parameters between solar cycles. In Chapter 3 we investigate Bayesian emulation as a possible alternative to genetic algorithms for parameter optimization. In Chapter 4 we simulate the evolution of individual real active regions from Cycles 21 to 24 using the 2D SFT model. We then calculate the end-of-cycle axial dipole moment contribution of each one in order to conclude whether a small number of regions can have a significant effect on the amplitude of the following cycle. For the simulations we determine optimal parameters using the same genetic algorithm as in Chapter 2. In Chapter 5 we develop a simple 2D radial diffusion model to help us understand why there is a discrepancy between the surface evolutions of the SFT model and a 3D kinematic dynamo model (Yeates & Muñoz-Jaramillo, 2013). The eventual goal is to optimize the parameters of the 3D model, but the surface magnetic field evolution problem should be corrected before any optimization can be performed. We conclude and discuss possible directions for future research in Chapter 6.

The work carried out in Chapters 2, 4 and 5 has been published in peer-reviewed journals (Whitbread *et al.*, 2017, 2018, 2019).

Chapter 2

Parameter optimization for surface flux transport models

Surface flux transport (SFT) is a crucial component of the 11-year sunspot cycle. SFT models have been developed and used since the 1960s with some success, though results can be sensitive to the choice of parameters. Parametrizations of the transport processes have been made, particularly for diffusion and meridional flow, but the exact forms are still debated, and those chosen are not always in line with the limited observations available. Parameter studies of varying scope have been undertaken (e.g., Schrijver *et al.*, 2002; Durrant & Wilson, 2003; Baumann *et al.*, 2004; Yeates, 2014; Virtanen *et al.*, 2017), but without complete parameter coverage. In this chapter we aim to systematically find optimal parameters to be used in SFT models so that they accurately reproduce such features of the solar cycle as poleward flux 'surges', polar field reversal time, polar field strength, and axial dipole moment. The results can also be used to constrain the surface components of dynamo simulations to produce the most accurate cycle predictions to date.

A similar study was performed recently by Lemerle *et al.* (2015), who used the same genetic algorithm used in this chapter to find optimal parameters for a 2D SFT model for Cycle 21 only, with the view of coupling it to a 2D flux transport dynamo model (Lemerle & Charbonneau, 2017). In contrast we analyse two distinct models

with different dimensionality, namely 1D and 2D, and different data-assimilation techniques, initially for Cycle 23. While Lemerle *et al.* (2015) used the Cycle 21 BMR database compiled by Wang & Sheeley (1989), we use the Cycle 23 BMR database from Yeates *et al.* (2007) for the 1D model, and extract individual active regions from synoptic magnetograms for the 2D model. This allows us to apply the 2D model to Cycles 21, 22, 23, and 24 to search for cyclical variation. For other studies of modelling Cycle 23 specifically, see e.g. Schrijver & Liu (2008), Yeates *et al.* (2010), Yeates & Mackay (2012) and Jiang *et al.* (2013).

In Section 2.1, we present the 1D model and the genetic algorithm used to perform the optimization, eventually including a prescribed error structure dependent on latitude and magnetic field strength in order to factor in observational uncertainty. We also discuss the results of the optimization for Cycle 23. In Section 2.2, we present the 2D model which directly assimilates active regions into the simulation, and run the optimization process on this model for Cycle 23. In Section 2.3, we compare our optimal meridional flow profiles from both models with observations. Finally, we perform optimizations on the 2D model for Cycles 21, 22, and 24 in Section 2.4, before concluding in Section 2.5. The results from this chapter have been published in Whitbread *et al.* (2017).

2.1 One-dimensional surface flux transport model

We begin with a 1D model because we would like to test the optimization algorithm on a simpler model that is quick to compute. Furthermore, because we will only optimize against longitude-averaged data, it is possible to use a 1D model in co-latitude only.

2.1.1 Derivation

In the absence of new flux emergence, the evolution of the radial component of the magnetic field is described by Equation 1.3.19. We now prescribe the velocity profiles u_θ and u_ϕ. In the azimuthal direction, the primary transport mechanism is differential rotation $u_\phi = \omega(\theta) R_\odot \sin\theta$ which (at the surface) depends only on co-latitude. For u_θ we have the meridional flow $v(\theta)$, which again only depends on co-latitude. Thus for $B \equiv B_r(\theta, \phi, t)$ we have:

$$\begin{aligned}\frac{\partial B}{\partial t} &= -\omega(\theta)\frac{\partial B}{\partial \phi} + \frac{\eta}{R_\odot^2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial B}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 B}{\partial\phi^2}\right] \\ &\quad - \frac{1}{R_\odot \sin\theta}\frac{\partial}{\partial\theta}\Big(v(\theta)\sin\theta\, B\Big) - \frac{1}{\tau}B + S(\theta,\phi,t), \end{aligned} \tag{2.1.1}$$

where η is turbulent diffusivity, and τ is the decay time for a hypothesised exponential decay term added by Schrijver *et al.* (2002) to improve regular polar field reversal. We include a source term S for newly emerging magnetic regions.

Since the equation is solved on a spherical surface, the magnetic field may be decomposed into spherical harmonic form (c.f. Baumann, 2005):

$$B(\theta,\phi,t) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm}(t)\, Y_{lm}(\theta,\phi), \tag{2.1.2}$$

where

$$Y_{lm}(\theta,\phi) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)\, e^{im\phi}, \tag{2.1.3}$$

with $P_l^m(\cos\theta)$ representing the associated Legendre polynomials. The coefficients $a_{lm}(t)$ are given by:

$$a_{lm}(t) = \int_0^{2\pi}\int_{-1}^{1} B(\theta,\phi,t)\, Y_{lm}^*(\theta,\phi)\, d(\cos\theta)\, d\phi. \tag{2.1.4}$$

Initially only one dimension, namely the latitudinal direction, is considered. We average the model spatially in the longitudinal dimension. It follows from averaging Equation 2.1.1 that differential rotation does not play a part in this model. The

surface flux transport equation then simplifies down to:

$$\frac{\partial \overline{B}}{\partial t} = \frac{\eta}{R_\odot^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial \overline{B}}{\partial\theta}\right) - \frac{1}{R_\odot \sin\theta}\frac{\partial}{\partial\theta}\left(v\left(\theta\right)\sin\theta\,\overline{B}\right) - \frac{1}{\tau}\overline{B} + S\left(\theta,t\right), \quad (2.1.5)$$

where $\overline{B}$ is the longitude-averaged radial field, and the magnetic field decomposition is also averaged over longitude:

$$\begin{aligned}
\overline{B}\left(\theta,t\right) &= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm}\left(t\right)\frac{1}{2\pi}\int_0^{2\pi} Y_{lm}\left(\theta,\phi\right)d\phi \\
&= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm}\left(t\right)\sqrt{\frac{\left(2l+1\right)}{4\pi}\frac{\left(l-m\right)!}{\left(l+m\right)!}}P_l^m\left(\cos\theta\right)\frac{1}{2\pi}\int_0^{2\pi} e^{im\phi}d\phi \\
&= \sum_{l=0}^{\infty} a_{l0}\left(t\right)\sqrt{\frac{2l+1}{4\pi}}P_l\left(\cos\theta\right), \qquad (2.1.6)
\end{aligned}$$

since $P_l^m\left(x\right) = \left(-1\right)^m\left(1-x^2\right)^{m/2}\dfrac{d^m}{dx^m}\left(P_l\left(x\right)\right)$, and we only need to consider the $m=0$ harmonic functions because the magnetic field does not depend on longitude. In Equation 2.1.5 we have subtly swapped the order of differentiation and integration. This is valid according to the Leibniz integral rule, because we have constant limits of integration and continuous B and $\overline{B}$, owing to the periodicity in the azimuthal direction.

Using the fact that the spherical harmonics are eigenfunctions of the Laplace operator, with eigenvalues $-l\left(l+1\right)$, and substituting the above form for the radial magnetic field into Equation 2.1.5, the following system of ordinary differential equations is obtained:

$$\begin{aligned}
\sum_{l=0}^{\infty}\dot{a}_{l0}\left(t\right)Y_{l0}\left(\theta\right) \;=\; & \sum_{l=0}^{\infty} -\frac{\eta l\left(l+1\right)}{R_\odot^2}a_{l0}\left(t\right)Y_{l0}\left(\theta\right) \\
& - \sum_{l=0}^{\infty}\frac{a_{l0}\left(t\right)}{R_\odot\sin\theta}\frac{\partial}{\partial\theta}\left(v\left(\theta\right)\sin\theta\,Y_{l0}\left(\theta\right)\right) \\
& - \sum_{l=0}^{\infty}\frac{a_{l0}\left(t\right)Y_{l0}\left(\theta\right)}{\tau} + \sum_{l=0}^{\infty} s_{l0}\left(t\right)Y_{l0}\left(\theta\right). \qquad (2.1.7)
\end{aligned}$$

Now using the orthogonality conditions

$$\int_{-1}^{1} P_l\left(\cos\theta\right)P_{l'}\left(\cos\theta\right)d\left(\cos\theta\right) = \frac{2}{2l+1}\delta_{ll'}, \qquad (2.1.8)$$

and

$$\int_0^{2\pi} e^{i(m-m')\phi} d\phi = 2\pi\delta_{mm'}, \tag{2.1.9}$$

Equation 2.1.7 can be multiplied by $Y^*_{l'0}(\theta)$ and integrated over $\int_0^{2\pi}\int_{-1}^{1} d(\cos\theta)\, d\phi$ to produce:

$$\begin{aligned}
\dot{a}_{l0}(t) &= -\frac{\eta l(l+1)}{R_\odot^2} a_{l0}(t) \\
&\quad - \frac{2\pi}{R_\odot}\sqrt{\frac{2l+1}{4\pi}} \sum_{l'=0}^{\infty} \sqrt{\frac{2l'+1}{4\pi}} a_{l'0}(t) \int_0^{\pi} \left[\frac{\partial v(\theta)}{\partial\theta} \sin\theta\, P_l(\cos\theta)\, P_{l'}(\cos\theta) \right. \\
&\quad + v(\theta)\cos\theta\, P_l(\cos\theta)\, P_{l'}(\cos\theta) \\
&\quad \left. + v(\theta)\sin\theta\, P_l(\cos\theta) \frac{\partial}{\partial\theta}\Big(P_{l'}(\cos\theta) \Big) \right] d\theta - \frac{a_{l0}(t)}{\tau} + s_{l0}(t).
\end{aligned} \tag{2.1.10}$$

Now using the recurrence relation for the derivative of a Legendre polynomial:

$$\frac{d}{dx}\big(P_l(x)\big) = \frac{l}{x^2-1}\big(xP_l(x) - P_{l-1}(x)\big), \tag{2.1.11}$$

Equation 2.1.10 becomes

$$\dot{a}_{l0}(t) = -\frac{\eta l(l+1)}{R_\odot^2} a_{l0}(t) - \frac{2\pi}{R_\odot}\sqrt{\frac{2l+1}{4\pi}} \sum_{l'=0}^{\infty} \sqrt{\frac{2l'+1}{4\pi}} a_{l'0}(t)\, C_{ll'} - \frac{a_{l0}(t)}{\tau} + s_{l0}(t), \tag{2.1.12}$$

where

$$\begin{aligned}
C_{ll'} &= \int_0^{\pi} \left[\frac{\partial v(\theta)}{\partial\theta} \sin\theta\, P_l(\cos\theta)\, P_{l'}(\cos\theta) \right. \\
&\quad + (l'+1)\, v(\theta)\cos\theta\, P_l(\cos\theta)\, P_{l'}(\cos\theta) \\
&\quad \left. - l' v(\theta)\, P_l(\cos\theta)\, P_{l'-1}(\cos\theta) \right] d\theta,
\end{aligned} \tag{2.1.13}$$

which can be solved numerically using MATLAB's inbuilt trapezoid rule solver. Equation 2.1.12 can then be written as a matrix equation:

$$\dot{a}_{l0}(t) = M_{ll'} a_{l'0}(t), \tag{2.1.14}$$

which we solve in MATLAB using the inbuilt explicit Runge-Kutta (4,5) time-stepping method (Dormand & Prince, 1980). The equations are solved on a grid of 180 cells equally spaced in latitude. The inbuilt solver combines a moderate accuracy

with an adaptive time step to automatically satisfy the Courant-Friedrichs-Lewy (CFL) condition (Courant *et al.*, 1928):

$$\frac{u\Delta t}{\Delta x} \leq C_{\max}, \tag{2.1.15}$$

where u is the magnitude of the velocity, Δt is the time step, and Δx is the spatial step. For explicit methods, $C_{\max}$ is typically set to unity. This ensures that our choice of grid spacing does not cause numerical instability. In addition, the stability of the explicit method relies on having the diffusion term. In an advection-only regime, we would need to use, for example, a finite-volume scheme.

The source terms of newly-emerging BMRs, $s_{l0}\,(t)$, are added into the equation on the corresponding day of emergence. In order to obtain these coefficients, the day of emergence (t_0); longitude (ϕ_0) and latitude (λ_0) of the BMR centre; half-separation distance between peaks of polarity (ρ_0); magnetic flux, including polarity (Φ_0); and tilt angle (γ_0) of each BMR must be given. The sources of these data will be described below.

After converting from spherical coordinates to Cartesian coordinates:

$$x = \sin\theta\,\cos\phi, \qquad y = \sin\theta\,\sin\phi, \qquad z = \cos\theta, \tag{2.1.16}$$

each BMR is rotated into the 'BMR frame' by the following matrix computation:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma_0 & -\sin\gamma_0 \\ 0 & \sin\gamma_0 & \cos\gamma_0 \end{pmatrix} \begin{pmatrix} \cos\lambda_0 & 0 & \sin\lambda_0 \\ 0 & 1 & 0 \\ -\sin\lambda_0 & 0 & \cos\lambda_0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{2.1.17}$$

Then after converting back to spherical coordinates in the BMR frame:

$$\lambda' = \delta\arcsin\left(z'\right), \qquad \phi' = \delta\arctan\left(\frac{y'}{x'}\right), \tag{2.1.18}$$

where $\delta = \dfrac{180}{\pi}$, the final expression for the BMR is given by

$$B\left(\theta,\phi,t\right) = -\frac{\Phi_0\,\delta^2\phi'}{\sqrt{\pi}\rho_0^3}\exp\left(-\frac{\left(\phi'\right)^2 + 2\left(\lambda'\right)^2}{2\rho_0^2}\right). \tag{2.1.19}$$

Since only one dimension is considered, this expression is averaged over ϕ by integrating via the trapezoid rule. The source term coefficients are calculated using Equation 2.1.4, i.e.,

$$s_{l0}(t) = \int_0^{\pi} \overline{B}(\theta, t)\, Y_{l0}(\theta) \sin\theta \, d\theta, \tag{2.1.20}$$

and are added onto the coefficients in Equation 2.1.14 at the time of emergence.

2.1.2 Optimization algorithm

To search for optimal parameter sets where the model matches the observed butterfly diagram, we use the genetic algorithm `PIKAIA 1.2`. It was written by Charbonneau & Knapp (1995) at the High Altitude Observatory (HAO) of the National Center for Atmospheric Research (NCAR) and is publicly accessible[a].

`PIKAIA` is an evolutionary algorithm written in FORTRAN-77 (though versions in other languages have been implemented). It is particularly efficient in multimodal optimization problems, taking advantage of a mutation operator which can induce random jumps in parameter space, while other optimization algorithms can get trapped at a local maximum and might fail to locate the global maximum.

The algorithm generates a set of parameters, each within a user-defined range, and runs a model simulation for each parameter set, or 'population member'. The population are ranked by 'fitness' according to a user-defined 'fitness function', which in our case will be a comparison between a real reference butterfly diagram and model-generated output.

The population members with higher ranks have a greater probability of being selected for the crossover or 'breeding' process whereby sections of the parameter strings are interchanged to produce 'offspring', in the hope that a fitter individual will be found with desirable features of both 'parents'. Random mutation of parameter string digits is included to increase variability and hence the likelihood for population improvement. In particular, the more recent versions of `PIKAIA` make use of a variable

[a] http://www.hao.ucar.edu/modeling/pikaia/pikaia.php

mutation rate – the rate increases as the population becomes denser and vice versa, allowing for a more efficient search of parameter space.

The breeding process is run over a pre-determined number of generations. Whilst `PIKAIA` is inherently stochastic and so convergence to a 'good' fit is never guaranteed, a large enough choice for the number of generations should ensure that the combined effect of crossover and mutation has enough time to introduce sufficiently fit population members. It should also be noted that the algorithm will almost certainly never find the exact optimal solution, but will converge towards it.

For a more in-depth introduction to the features and operation of `PIKAIA` and genetic algorithms in general, see Charbonneau (2002*a*) and Charbonneau (2002*b*).

2.1.3 Testing `PIKAIA` on artificial data

To test the success and efficiency of `PIKAIA`, the algorithm is used on the 1D SFT model from Section 2.1.1 with 5474 randomly generated artificial BMRs emerging over the course of three cycles. The simulations are compared to a reference case solution created using prescribed flows and fixed parameters, which the algorithm must successfully match. The synthetic cycle is created by drawing on statistics of real solar cycles, generating the day of emergence, longitude and latitude of the BMR centre, half-separation distance between peaks of polarity, magnetic flux and tilt angle for each artificial BMR. The BMRs are then converted into the BMR frame in the same way as described in Section 2.1.1.

Meridional flow is modelled using the Schüssler-Baumann velocity profile (Schüssler & Baumann, 2006), which is a combination of exponential and sinusoidal functions adapted to helioseismic observations (Gizon & Duvall, 2004):

$$v(\theta) = -v_0 \sin(2\theta) \exp\left(\pi - 2\left|\tfrac{\pi}{2} - \theta\right|\right), \tag{2.1.21}$$

where v_0 is chosen to set the maximum velocity; in this case $v_0 = 16\,\mathrm{m\,s^{-1}}$. In order to define the initial conditions, we adopt a simple sinusoidal profile (e.g. Svalgaard

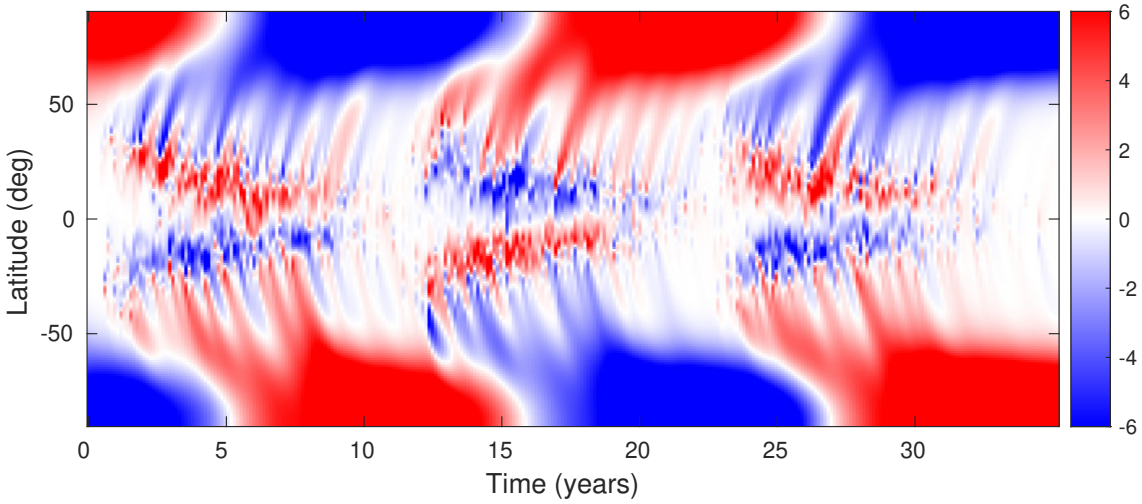

Figure 2.1: Reference case butterfly diagram generated from artificial BMRs.

et al., 1978):

$$\overline{B}\,(\theta, 0) = B_0 |\cos\theta|^7 \cos\theta, \tag{2.1.22}$$

with $B_0 = 10\,\mathrm{G}$. Other chosen parameters include $\eta = 500\,\mathrm{km^2\,s^{-1}}$ and $\tau = 10\,\mathrm{yr}$, and l runs from 1 to 128; the monopole term $l = 0$ is omitted. The butterfly diagram produced from these parameters is shown in Figure 2.1. The diagram suggests that the model successfully reproduces realistic features such as equatorward migration of active region emergence, poleward transport of flux and antisymmetry of parity about the equator.

An initial diagnostic test is run to test whether `PIKAIA` can select the correct diffusion while keeping all other parameters fixed as above, including the synthetic BMRs. A physically plausible range of $100\,\mathrm{km^2\,s^{-1}} \leq \eta \leq 1500\,\mathrm{km^2\,s^{-1}}$ is given to `PIKAIA` to search between. Twenty population members are generated and the breeding process runs over 32 generations, totalling 640 model simulations. For a model that takes about 90 seconds to run, the total runtime for the initial diagnostic optimization comes to about 16 hours. Other parameters within `PIKAIA` are set to their default values: parameter strings have length 6, probability of crossover is 0.85, and initial, minimum and maximum values of the mutation rate are 0.005, 0.001 and $\frac{1}{6k}$ respectively, where k is the number of parameters.

A reduced χ^2 is used as a fitness function:

$$\chi^2 = \sqrt{\frac{1}{n-k}\sum_{i,j}\left(\overline{B}_{\rm map}\left(\theta_i, t_j\right) - \overline{B}\left(\theta_i, t_j; \mathbf{X}\right)\right)^2}, \tag{2.1.23}$$

where n is the number of gridpoints, $\mathbf{X}$ is the set of k parameters and $(n-k)^{-1}$ is a normalization factor: the number of degrees of freedom of the fit. Since a smaller value of χ^2 indicates a better fit, and `PIKAIA` operates by finding a maximum fitness value, the final fitness function is χ^{-2}, the reciprocal of Equation 2.1.23.

The diagnostic run returns an optimal value of $\eta = 501.8\,\mathrm{km^2\,s^{-1}}$. As noted in Section 2.1.2, the exact solution is never likely to be found, especially with time constraints, but an error at less than 0.5% is very encouraging. The corresponding butterfly diagram is qualitatively identical to Figure 2.1.

Increasing the number of parameters for optimization slows the rate of convergence, yet still produces qualitatively successful results (Table 2.1). The maximum velocity is given a range of $5\,\mathrm{m\,s^{-1}} \leq v_0 \leq 30\,\mathrm{m\,s^{-1}}$, and the decay time is restricted to $0\,\mathrm{yr} \leq \tau \leq 32\,\mathrm{yr}$. One notable aspect of the results is the sensitive balance between accuracy and work; whilst the optimal parameters for $n = 2$ produce an excellent fit, it should be noted that many more population members and breeding generations are used, culminating in a runtime of at least 65 hours. Any meaningful increase on these factors would make the problem computationally too expensive. The results for $n = 3$ highlight the first significant error in the optimization process, namely a deficiency of $36.3\,\mathrm{km^2\,s^{-1}}$ in η, owing to the combination of a shorter runtime and more free parameters. Each of the parameter sets in Table 2.1 produce butterfly diagrams that are qualitatively identical to Figure 2.1. The differences in values of np and ng are due to restrictions on available time.

Next, a more flexible meridional flow profile is chosen to replicate the situation of working with real data where the flow profile is not precisely known, as mentioned in Section 1.2.1:

$$v\left(\theta\right) = -v_0 \sin^p\theta\cos\theta, \tag{2.1.24}$$

n	np	ng	χ^{-2}	η ($\mathrm{km^2\,s^{-1}}$)	v_0 ($\mathrm{m\,s^{-1}}$)	τ (yr)
Reference	-	-	-	500.0	16.0	10.0
1	20	32	100.0	501.8	-	-
2	40	65	561.0	499.7	16.0	-
3	30	50	5.68	463.7	15.0	10.1

Table 2.1: Optimization outputs from `PIKAIA` runs for increasing numbers of parameters. The number of parameters is n, the population size is np, and ng is the number of generations.

where p is a free parameter left to be optimized ($0 \leq p \leq 16$). Increasing p produces a steeper gradient at low latitudes and a peak closer to the equator (Figure 2.2). Again, v_0 is chosen to be the maximum of $|v|$, so increasing v_0 increases the height of the peak. This severely restricts the optimization process, as shown in Table 2.2. While more flexible in theory than setting a defined function as in Equation 2.1.21, this form is unable to match the exact shape of the reference-case flow and so the fitness is substantially smaller, and maximum velocities larger. Figure 2.3 shows a direct comparison between the reference case flow and the optimal `PIKAIA`-generated flow. Whilst the model flow matches the reference case effectively at the equator, it performs poorly in the upper half of the activity belts (15°–30°) and the poleward-surge flux transport regions (30°–50°), before providing a closer fit at the poles. This could be due to flux-cancellation at the equator being an important factor in determining the polar field strength for that cycle, since it is the uncancelled trailing flux that is transported polewards to form the new polar field.

In order to reduce the duration of computationally intensive optimizations, Metcalfe & Charbonneau (2003) created `MPI`*`KAIA`*, a freely accessible implementation of `PIKAIA` 1.2 in MPI[b]. Rather than using a single processor for all model evaluations, a network of computers is used, and each of the model evaluations from a single generation is sent to a separate processor and computed simultaneously, achieving near-perfect parallelization. The time taken to complete the optimization therefore

[b]http://www.hao.ucar.edu/Public/about/Staff/travis/mpikaia/

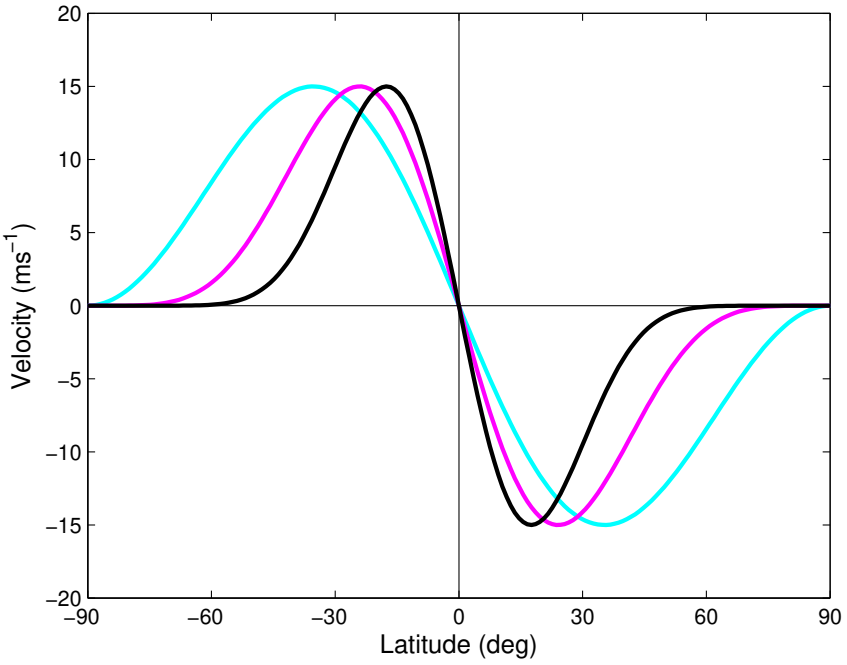

Figure 2.2: Three examples of the meridional flow profile in Equation 2.1.24 for $v_0 = 15\,\mathrm{m\,s^{-1}}$, $p = 2$ (cyan), $p = 5$ (magenta) and $p = 10$ (black).

n	np	ng	χ^{-2}	η ($\mathrm{km^2\,s^{-1}}$)	v_0 ($\mathrm{m\,s^{-1}}$)	p	τ (yr)
Reference	-	-	-	500.0	16.0	-	10.0
2	18	35	1.05	-	21.0	3.56	-
4	20	100	1.31	431.2	19.7	3.31	11.2

Table 2.2: Optimization outputs from `PIKAIA` runs including the flexible meridional flow profile in Equation 2.1.24. The number of parameters is n, the population size is np, and ng is the number of generations.

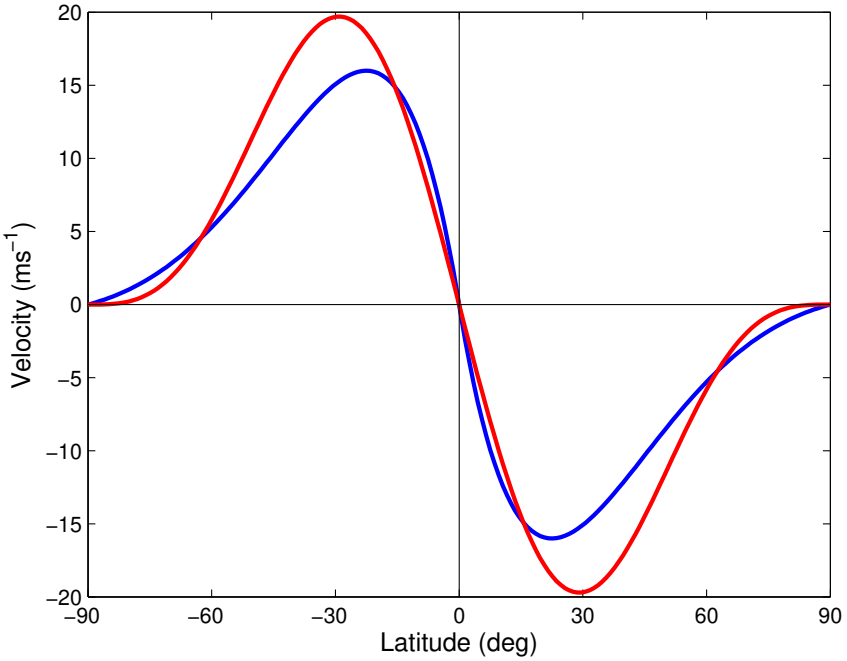

Figure 2.3: Comparison of the reference case meridional flow profile (blue) and the profile generated using the parameters from the $n = 4$ case in Table 2.2 (red).

is entirely dependent on the number of generations. For example, for the 1D optimization problem in this section, each model simulation takes approximately a minute and a half, so, with 46 processors, the runtime is brought down from 24 days to about 12.5 hours for 46 population members evolved over 500 generations. These choices of population size and evolution period should be large enough to obtain good fits to the data, and will be used for the remainder of the study unless specified.

With `MPIKAIA` set up, more extensive optimization runs can be processed, the results of which are shown in Table 2.3. Note that a fifth parameter B_0, the maximum initial polar field strength, is eventually included, with a prescribed range of $0\,\mathrm{G} \leq B_0 \leq 25\,\mathrm{G}$. Whilst the optimization is restricted for $n = 2$ since all other parameters are fixed, including two more variables in the process allows the algorithm to search for a better overall fit by altering these new variables. This is evident from the results in Table 2.3, where the latter three parameter sets give higher values of χ^{-2} than the $n = 2$ case. This behaviour of increasing or decreasing parameters to accommodate additional ones also explains the reduction of $\sim 100\,\mathrm{km^2\,s^{-1}}$ in η.

n	np	ng	χ^{-2}	η ($\mathrm{km^2\,s^{-1}}$)	v_0 ($\mathrm{m\,s^{-1}}$)	p	τ (yr)	B_0 (G)
Reference	-	-	-	500.0	16.0	-	10.0	10.0
2	30	250	1.18	-	20.9	3.32	-	-
4	30	200	1.32	399.1	18.9	3.28	10.8	-
4	46	500	1.33	405.7	19.0	3.25	10.7	-
5	46	500	1.32	416.4	18.8	3.22	10.0	11.0

Table 2.3: Optimization outputs from `MPIKAIA` runs for increasing numbers of parameters. The number of parameters is n, the population size is np, and ng is the number of generations.

The resulting butterfly diagram arising from the optimal 5-parameter set in Table 2.3 is presented in the top panel of Figure 2.4, with the reference butterfly diagram in the bottom panel for direct qualitative comparison. The model reconstructs the solution reasonably accurately, particularly in emergence regions near the equator. The field at the poles is too strong owing to the larger maxima of the flexible velocity profiles. These maxima are usually located around the bottom of the transport regions, meaning more trailing flux is quickly transported polewards resulting in a stronger polar field.

2.1.4 Cycle 23 analysis

With both the algorithm and model sufficiently tested, our attention turns to using the same process to optimize the model for Solar Cycle 23. Now each BMR has a specified day of emergence; longitude and latitude; size; magnetic flux, including polarity; and tilt angle taken from an existing observational dataset where these properties were determined individually for each BMR from National Solar Observatory (NSO) synoptic magnetograms (Yeates *et al.*, 2007). Using these data, the 1644 recorded BMRs from Cycle 23 (1996 June 1–2008 August 3) are converted into the appropriate frame using the same procedure described in Section 2.1.1, and inserted into the model on the corresponding days of emergence. The BMR data are freely

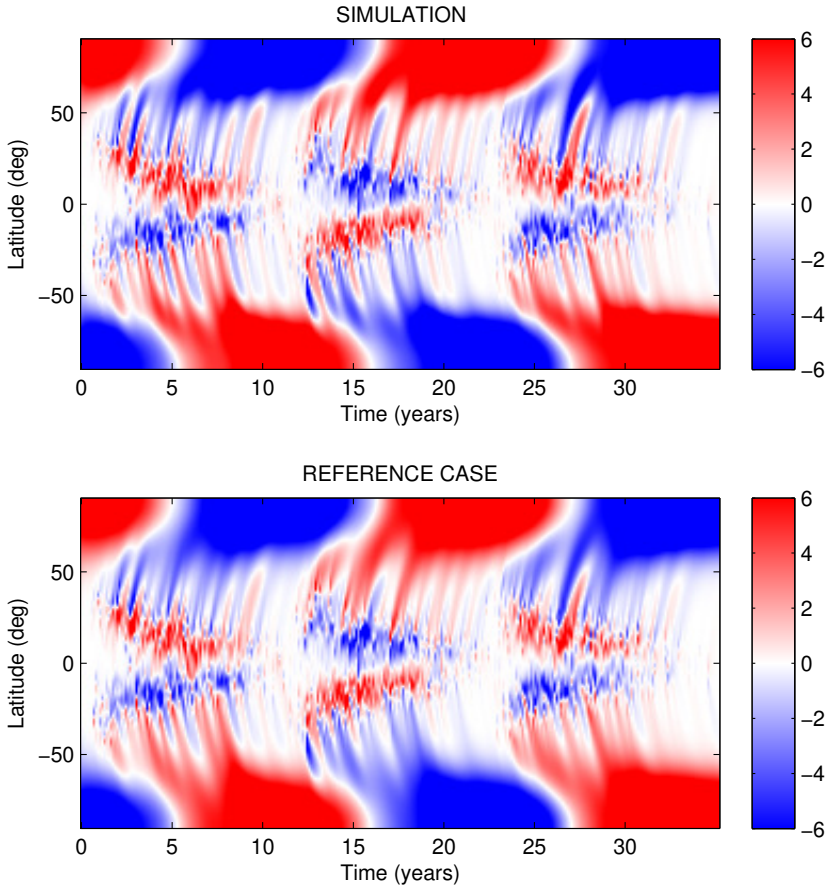

Figure 2.4: Butterfly diagram produced from the optimal 5-parameter set in Table 2.3.

available at the Solar Dynamo Dataverse[c] (Yeates, 2016). As shown in Figure 2.5, the database includes regions of each of the four tilt/polarity combinations. Of the 1644 BMRs from the database, 1127 are positive tilt and normal polarity (Case I), 38 are positive tilt and abnormal polarity (Case II), 39 are negative tilt and abnormal polarity (Case III), and 440 are negative tilt and normal polarity (Case IV). Note that the low-latitude Case II/III active regions at the start of Cycle 23 are more likely to be Case I/IV regions from Cycle 22, and similarly with the high-latitude regions at the end of Cycle 23 actually belonging to Cycle 24. These are included in the classification of Cycle 23 because of overlapping cycles. However, they should not confuse the statistics of relative frequency of different active region orientation cases.

The same functional form for the meridional flow in Equation 2.1.24 is used. Lemerle *et al.* (2015) used a similar, but more sophisticated profile, which is discussed in Section 2.4.1. This provides substantially more flexibility, but introduces extra parameters into the optimization runs which could hinder convergence to a global maximum. Conversely, van Ballegooijen *et al.* (1998) used a basic sinusoidal profile which stopped abruptly at $\pm 75°$. While a simpler functional form in practice, this does not provide any flexibility to aid the optimization. In any case, the true functional form of the observed meridional circulation is uncertain, particularly at high latitudes.

Figure 2.6 (blue) shows the observed initial $\overline{B}$ profile from June 1996. The profile is asymmetric across the equator in terms of polar field strength, and there is some activity present at the equator. The curve given by Equation 2.1.22 and used in the simulations is shown in red in Figure 2.6. This represents a typical cycle minimum profile and ensures that the choice of initial profile is not hindering the optimization process, but rather aiding it with some flexibility. We do not allow for asymmetry because it is difficult to parametrize and vary using the algorithm whilst ensuring flux conservation.

[c]https://dataverse.harvard.edu/dataverse/solardynamo

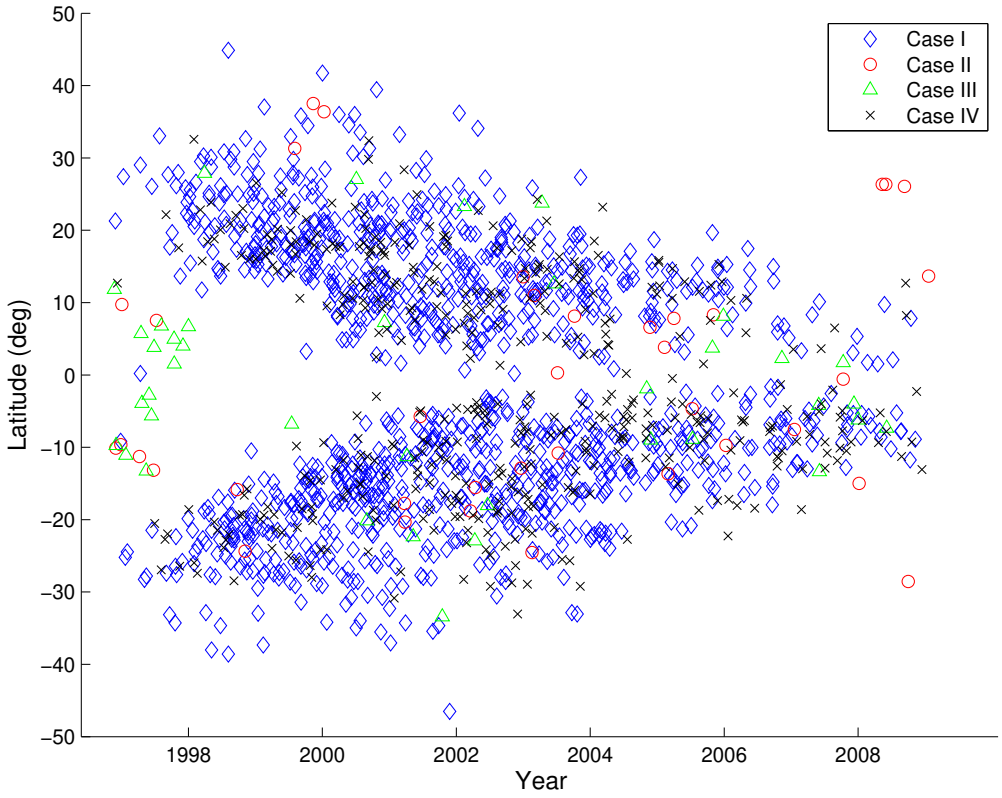

Figure 2.5: Scatterplot of the four cases of BMR alignment. Case I is positive tilt and normal polarity, Case II is positive tilt and abnormal polarity, Case III is negative tilt and abnormal polarity, and Case IV is negative tilt and normal polarity. Normal (abnormal) refers to when the leading polarity matches (opposes) the typical leading polarity of the cycle in question. Positive (negative) tilt refers to when the leading polarity is closer to (further from) the equator than the trailing polarity.

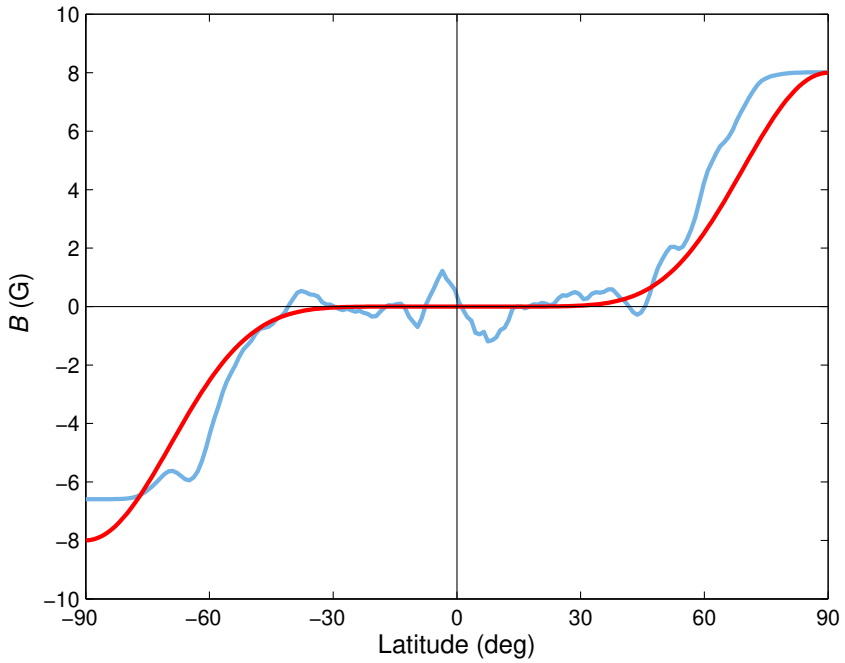

Figure 2.6: Comparison between initial magnetogram (blue) and the profile given in Equation 2.1.22 (red) with $B_0 = 8\,\mathrm{G}$.

2.1.5 Ground-truth data

As ground-truth data for optimization of the model, we use radial-component magnetogram data from US National Solar Observatory, Kitt Peak, in the form of full-disk images. Prior to Carrington rotation (CR) 2007, these came from the Kitt Peak Vacuum Telescope, while 2007 CR onwards we use Synoptic Optical Long-term Investigations of the Sun (SOLIS) data[d]. The original magnetograms combine to produce a synoptic map with noise in the polar regions and data gaps for particular Carrington rotations, visible in the top panel of Figure 2.7. To minimize noise, Petrie (2012) corrected the butterfly diagram by calculating a cubic spline interpolation at each latitude of annual average measurements of high-latitude fields (poleward of $\pm 75°$) which were observed with a preferable solar rotation axis tilt angle. A combination of real and interpolated data was used for the regions between $\pm 60°$ and $\pm 75°$. The resulting butterfly diagram is interpolated onto a uniform time grid at daily intervals. This is averaged over periods of 27 days, smoothed using a Gaussian filter

[d]http://solis.nso.edu/0/vsm/vsm_maps.php

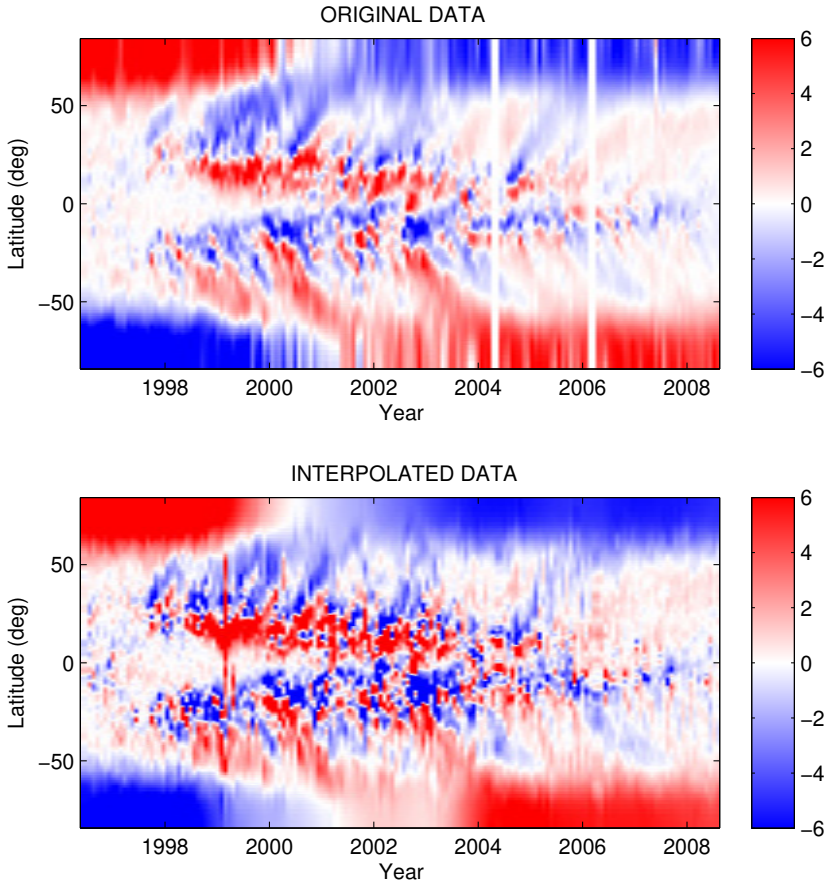

Figure 2.7: Top: Original Kitt Peak magnetogram data for Cycle 23. Bottom: Interpolated data to be used as 'ground truth'.

to bring the unsigned flux down to a comparable level to the simulation, and finally sampled at the resolution of once per Carrington rotation (27.2753 days), as shown in the bottom panel of Figure 2.7. Unfortunately, the correction process results in a non-physical spike in the data around 1999. While not ideal, only a small proportion of the image is affected, so the genetic algorithm will not be severely hindered by the error.

2.1.6 Initial results

We begin with the same five parameters as in Table 2.3 to be optimized initially. Maximum and minimum limits are similar to those in Section 2.1.3 and are prescribed based on results from literature and observations (e.g. Schrijver *et al.*, 2002; Hathaway

& Rightmire, 2010; Yeates, 2014; Lemerle *et al.*, 2015):

(i) $100\,\mathrm{km^2\,s^{-1}} \leq \eta \leq 1500\,\mathrm{km^2\,s^{-1}}$

(ii) $5\,\mathrm{m\,s^{-1}} \leq v_0 \leq 30\,\mathrm{m\,s^{-1}}$

(iii) $0 \leq p \leq 16$

(iv) $0\,\mathrm{yr} \leq \tau \leq 32\,\mathrm{yr}$

(v) $0\,\mathrm{G} \leq B_0 \leq 50\,\mathrm{G}$

It should be noted that these ranges are deliberately made wider than results from literature to allow for a deeper exploration into the parameter space and to provide a better understanding of the SFT model. Table 2.4(a) shows the results of the 1D optimization for Cycle 23. Other entries in Table 2.4 show various tests which are introduced in later sections.

During an optimization run, every single population member generated by `PIKAIA` can be recorded, and so a range of 'acceptable' values can be obtained for each parameter. These can be found in square brackets below each optimal value in Table 2.4. The upper and lower bounds are taken to be the largest and smallest values for each parameter which produce fits above 95% of the maximum χ^{-2}. Anything within these limits is classed as 'acceptable', though it must be noted that choosing to fix one parameter can alter the optimal solutions and bounds for others.

However, the values of χ^{-2} in Table 2.4 come into question since they have no real meaning. Thus far we have not taken into account model error or observation error, and we have assumed that each data point is independent. Ultimately this does not allow for comparison of separate regimes, and we cannot make statistically meaningful conclusions from the results. This was not an issue in Section 2.1.3, because there was no 'error' in the model or the 'ground truth data' per se. The data were produced from the model, and so it was theoretically possible in that case to achieve a perfect match (leading to an infinite χ^{-2}). Consequently we now consider

Regime	χ^{-2}	η ($\mathrm{km^2\,s^{-1}}$)	v_0 ($\mathrm{m\,s^{-1}}$)	p	τ (yr)	B_0 (G)	γ	B_{par} (G)
				Cycle 23				
(a) 1D	0.45	330.5 [186.3, 624.4]	14.0 [9.0, 28.5]	3.19 [2.64, 4.72]	2.7 [2.0, 5.2]	15.1 [10.5, 20.1]	1.00	n/a
(b) 1D + γ	0.50	356.0 [212.5, 662.7]	11.1 [7.3, 20.1]	2.42 [1.95, 3.32]	3.6 [2.7, 6.4]	11.8 [8.9, 15.2]	0.57 [0.42, 0.74]	n/a
(c) 2D + B_{par}	0.40	440.4 [273.1, 768.2]	10.7 [7.4, 14.9]	2.82 [1.83, 5.27]	n/a	8.2 [5.1, 11.8]	n/a	39.4 [21.1, 48.7]
(d) 2D + τ	0.40	426.8 [249.9, 852.5]	9.1 [5.1, 14.2]	2.23 [1.61, 4.14]	5.0 [1.8, 31.9]	11.6 [5.8, 18.7]	n/a	39.4
(e) 1D, fixed p	0.44	315.8 [161.7, 630.9]	8.0 [5.2, 11.0]	1.87	2.1 [1.6, 2.9]	14.5 [9.5, 19.8]	1.00	n/a
(f) 2D, fixed p	0.38	465.7 [289.3, 831.0]	11.9 [6.6, 16.9]	1.87	n/a	9.4 [6.4, 13.9]	n/a	39.4
				Cycle 21				
(g) 2D	0.38	455.7 [301.5, 889.7]	9.5 [5.5, 14.5]	2.49 [1.68, 4.29]	n/a	6.2 [3.2, 9.8]	n/a	39.4
				Cycle 22				
(h) 2D	0.37	479.0 [289.0, 889.2]	8.4 [5.2, 15.9]	2.47 [1.67, 4.08]	n/a	9.4 [5.5, 11.9]	n/a	39.4
				Cycle 24				
(i) 2D	0.55	453.5 [242.1, 1160.4]	8.1 [5.1, 15.6]	2.17 [0.57, 14.81]	n/a	4.1 [2.0, 6.1]	n/a	39.4

Table 2.4: Optimal parameter sets for each optimization regime. Underlined entries represent parameters that were fixed for the corresponding run. Upper and lower bounds for acceptable parameter ranges are given in square brackets below each entry.

an observation-based error structure which is discussed below in Section 2.1.7. We find that including variance in the fitness calculation constrains the optimization further by decreasing the range of acceptable parameter values. Moreover, it makes minimal difference to the optimal values (see Section 2.1.8), and so we will continue to use this updated fitness function unless stated otherwise.

2.1.7 Accounting for uncertainties in observations

We now use an improved χ^2 statistic as a measure of fit between the real and simulated butterfly diagrams:

$$\chi^2 = \frac{1}{n-k} \sum_{i,j} \left(\frac{\overline{B}_{\text{obs}}(\theta_i, t_j) - \overline{B}(\theta_i, t_j; \mathbf{X})}{\sigma(\theta_i, t_j)} \right)^2, \tag{2.1.25}$$

where n is the number of gridpoints and $\mathbf{X}$ is the vector of k free parameters. Again, since improving best fit is a minimization process and `PIKAIA` is set up to maximize functions, the reciprocal of the measure, χ^{-2}, is used as the necessary fitness function. The variance σ^2 describes the error in both the measurements and the models, and we assume the form:

$$\sigma^2(\theta_i, t_j) = \sigma^2_{\text{obs}}(\theta_i, t_j) + \sigma^2_{\text{model}}. \tag{2.1.26}$$

This is because the observational and model errors are independent, and other combinations would not allocate an appropriate error to each pixel. The variance plays two roles in the optimization. Firstly, it gives a meaningful value to the χ^{-2} statistic. This allows us to compare the performance of different parameter combinations and time periods. Secondly, it effectively weights distinct locations (θ_i, t_j) differently in the optimization, since the observed errors are assumed to have the form:

$$\sigma_{\text{obs}}(\theta_i, t_j) = \frac{0.1 \left| \overline{B}_{\text{obs}}(\theta_i, t_j) \right| + \varepsilon}{\sin \theta_i}, \tag{2.1.27}$$

where ε is some small increment to ensure that the error is non-zero even in regions of weak magnetic field. This error structure reflects the uncertainties and inconsistencies in photospheric magnetic field observations (e.g. Riley *et al.*, 2014). The factor of

$\sin\theta$ allows for the fact that the errors are in the original line-of-sight measurements, whereas $\overline{B}_{\rm obs}$ is the inferred radial field (Svalgaard *et al.*, 1978). Overall, the effect of this error structure reduces the weight of observations both near the pole and in strong active regions, as well as non-physical spikes as discussed in Section 2.1.5. The resulting optimization will favour accuracy in the mid-latitude 'transport regions'.

Strictly speaking, we should construct a covariance matrix:

$$\mathrm{Cov}\left(\overline{B}\left(\mathbf{x}\right),\overline{B}\left(\mathbf{x}'\right)\right)=\sigma\left(\mathbf{x}\right)\sigma\left(\mathbf{x}'\right)\exp\left(-\sum_{i=1}^{N}\left(\frac{x_i-x_i'}{\Theta_i}\right)^2\right), \tag{2.1.28}$$

where N is the number of dimensions of the dataset (here $N=2$), and Θ_i is some correlation length for each dimension. Then we could calculate χ^2 as:

$$\chi^2=\frac{1}{n-k}\left(\overline{\mathbf{B}}_{\rm obs}-\overline{\mathbf{B}}\right)\,\mathrm{Cov}^{-1}\,\left(\overline{\mathbf{B}}_{\rm obs}-\overline{\mathbf{B}}\right)^T. \tag{2.1.29}$$

However, when correlation lengths are included, we find that the algorithm accounts for both error and correlation by increasing the diffusivity η in order to smooth out the simulated butterfly diagram. This pushes the resulting optimal value of η to the upper limit of exploration, high above any value that has been used or estimated before in such models. For this reason and for simplicity, we ignore correlation in this study and assume independence. The covariance matrix then only contains diagonal terms, and the χ^2 calculation reduces to Equation 2.1.25.

Since the model error structure is unknown, we compute χ^{-2} with $\sigma_{\rm model}=0$. This is sufficient for the purpose of comparing different model runs against the same set of observational data, with a higher value of χ^{-2} indicating models that give a better match. The simulations are not sufficiently detailed to achieve a significant match at, say, the 99% level, which is evident from visual inspection of the butterfly diagrams. To achieve such a close match would be very challenging, since the large numbers of degrees of freedom $n-k\sim16\,000$–$30\,000$ mean that the 99% interval for the χ^{-2} statistic is narrow, typically $[0.98,1.02]$.

In principle, we could estimate $\sigma_{\rm model}$ by increasing it and broadening the 99% interval until the value of χ^{-2} falls within this interval. This would give a meaningful estimate

of the 'model error' in a particular run. But this would not change the ordering of different model runs, or indeed the final optimal parameters, so we have not included such analysis here. For a comprehensive overview of multivariate analysis, see Mardia *et al.* (1979), and for more detailed approaches to dealing with model error, see Goldstein *et al.* (2013).

2.1.8 Error-weighted results

Optimal parameters

The optimization runs are performed again in the same way as above, but with the addition of σ in the calculation of the fitness value. Table 2.5(a) shows the results of the 1D optimization for Cycle 23, with the corresponding optimal butterfly diagram in the top panel of Figure 2.8, and the interpolated NSO data for Cycle 23 discussed in Section 2.1.5 in the bottom panel.

The equatorward migration of active regions is well represented by the BMR data, and large poleward surges are reproduced by the model. While the southern polar field reversal is well approximated by the model, the reversal in the northern hemisphere has a delay of approximately 6 months. Furthermore, there are multiple weak poleward surges in the simulated butterfly diagram, most noticeably around 2004, which do not appear in the real butterfly diagram. This is likely to be a by-product of approximating regions as BMRs and overestimating the contribution of flux from smaller regions. This build-up of flux results in a strong polar field that extends to lower latitudes, requiring a short decay timescale as is found in the optimization.

Parameter analysis

Acceptable parameter ranges introduced in Section 2.1.6 can be found in square brackets below each optimal value in Table 2.5. Figure 2.9 shows such bounds, denoted by the left and right vertical purple lines on each plot, for all parameter

Regime	χ^{-2}	η ($\mathrm{km^2\,s^{-1}}$)	v_0 ($\mathrm{m\,s^{-1}}$)	p	τ (yr)	B_0 (G)	γ	B_{par} (G)
				Cycle 23				
(a) 1D	0.89	351.6 [229.4, 546.9]	14.0 [11.3, 22.5]	3.24 [2.98, 4.50]	2.4 [1.9, 3.5]	16.5 [12.8, 20.9]	1.00	n/a
(b) 1D + γ	1.09	373.5 [233.3, 582.8]	11.0 [8.2, 16.5]	2.44 [1.82, 2.96]	3.7 [2.9, 6.1]	11.7 [8.5, 14.8]	0.55 [0.41, 0.61]	n/a
(c) 2D + B_{par}	0.65	455.6 [371.6, 651.0]	11.2 [8.6, 14.4]	2.76 [1.64, 4.71]	n/a	8.3 [5.3, 10.2]	n/a	39.8 [31.7, 49.4]
(d) 2D + τ	0.67	453.5 [299.5, 807.7]	9.6 [6.8, 15.2]	2.15 [1.50, 3.95]	4.5 [1.6, 30.3]	12.9 [6.5, 18.4]	n/a	39.8
(e) 1D, fixed p	0.85	361.4 [220.1, 642.8]	8.3 [7.4, 10.9]	1.87	1.9 [1.5, 2.3]	16.3 [11.8, 21.4]	1.00	n/a
(f) 2D, fixed p	0.64	482.1 [356.1, 712.9]	11.5 [8.8, 15.2]	1.87	n/a	9.7 [7.1, 12.8]	n/a	39.8
(g) 2D, half cycle	0.48	482.3 [315.1, 779.0]	11.2 [7.9, 14.4]	2.60 [1.64, 5.40]	n/a	8.3 [5.0, 11.3]	n/a	39.8
(h) 1D, observed initial profile	0.81	294.7 [195.1, 476.0]	13.4 [10.0, 25.5]	3.28 [3.06, 5.04]	2.5 [1.9, 4.6]	n/a	1.00	n/a
(i) 2D, observed initial profile	0.65	499.0 [353.0, 663.3]	11.1 [8.3, 14.0]	2.51 [1.61, 4.21]	n/a	n/a	n/a	39.4 [29.4, 47.8]
				Cycle 21				
(j) 2D	0.87	455.7 [342.7, 667.0]	9.2 [6.6, 12.0]	2.33 [1.33, 3.93]	n/a	6.6 [4.5, 9.4]	n/a	39.8
(k) 2D + τ	0.88	454.7 [303.8, 726.1]	8.5 [5.7, 11.5]	2.04 [0.96, 3.64]	10.2 [3.1, 32.0]	7.6 [3.6, 12.6]	n/a	39.8
				Cycle 22				
(l) 2D	0.84	506.2 [365.1, 760.9]	8.7 [6.1, 11.7]	2.18 [0.98, 3.60]	n/a	10.5 [7.5, 13.8]	n/a	39.8
(m) 2D + τ	0.86	505.7 [351.8, 838.5]	7.9 [5.3, 12.1]	1.93 [1.16, 3.53]	7.6 [3.1, 32.0]	12.8 [7.4, 18.2]	n/a	39.8
				Cycle 24				
(n) 2D	0.99	454.6 [292.6, 821.7]	8.2 [5.4, 12.5]	2.05 [0.62, 5.22]	n/a	4.2 [2.6, 5.4]	n/a	39.8
(o) 2D + τ	0.99	453.5 [268.0, 829.0]	8.0 [5.0, 12.9]	1.98 [0.73, 5.18]	15.1 [2.5, 32.0]	5.0 [3.0, 7.5]	n/a	39.8

Table 2.5: Optimal parameter sets for each optimization regime. Underlined entries represent parameters that were fixed for the corresponding run. Upper and lower bounds for acceptable parameter ranges are given in square brackets below each entry. The ranges for regime (a) are presented visually in Figure 2.9.

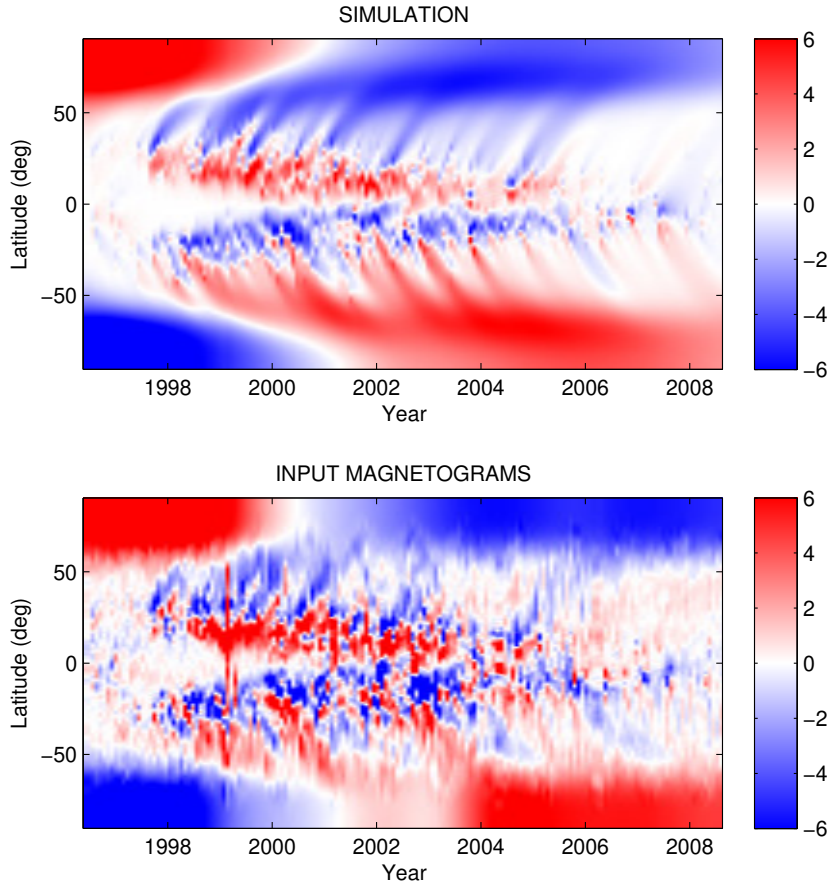

Figure 2.8: Top: Butterfly diagram for the optimal parameter 5-set for the 1D model in Table 2.5(a). Bottom: Ground truth data for Cycle 23.

populations from the optimization run that produced the optimal set in Table 2.5(a). The optimal values are highlighted by the central vertical purple lines. Using the limits for v_0 and p, acceptable meridional flow profiles are also found which are represented by the purple shading in the bottom right panel. The bold purple profile in the same panel represents the optimum profile.

The diffusion parameter η has not yet been accurately measured, though some indirect measurements by Mosher (1977) and Komm *et al.* (1995) have found values within the range of 100–300 $\mathrm{km^2\,s^{-1}}$. Early simulations by Leighton (1964) used values up to 1000 $\mathrm{km^2\,s^{-1}}$, though studies by Baumann *et al.* (2004), Wang *et al.* (1989*b*) and Wang & Sheeley (1991) decreased it to $\sim 600\,\mathrm{km^2\,s^{-1}}$, before Wang *et al.* (2002*b*) reduced it further to 500 $\mathrm{km^2\,s^{-1}}$. Our optimal value of 351.6 $\mathrm{km^2\,s^{-1}}$, however, is in better agreement with Yeates (2014), who found that $\eta \in [200, 450]\,\mathrm{km^2\,s^{-1}}$ produced a reasonable correlation between the butterfly diagrams, and Lemerle *et al.* (2015) who found an optimal value of 350 $\mathrm{km^2\,s^{-1}}$ within an acceptable range of 240–660 $\mathrm{km^2\,s^{-1}}$ for Cycle 21. Furthermore, Schrijver (2001) and Thibault *et al.* (2014) found diffusion coefficients of 300 $\mathrm{km^2\,s^{-1}}$ and 416 $\mathrm{km^2\,s^{-1}}$ respectively for random-walk-based models, and Cameron *et al.* (2016) recently used a diffusion of 250 $\mathrm{km^2\,s^{-1}}$. The acceptable range in Table 2.5(a) is broad but can be attributed to multiple degrees of freedom in the optimization. The range covers most values discussed above.

The large-scale meridional flow is poorly constrained by observations, as discussed in Section 1.2.1. Nevertheless, our optimal value of $v_0 = 14\,\mathrm{m\,s^{-1}}$ is in accordance with both the observations and simulations. Doppler measurements by Ulrich (2010) estimated the maximum velocity to be between 14–16 $\mathrm{m\,s^{-1}}$ for Cycles 22 and 23. Hathaway & Rightmire (2010) obtained an average maximum velocity of 10–12 $\mathrm{m\,s^{-1}}$ for Cycle 23 via magnetic feature tracking, though crucially they observed that the flow is slower (approximately 8 $\mathrm{m\,s^{-1}}$) at cycle maximum and faster (11.5–13 $\mathrm{m\,s^{-1}}$) at minimum. This time-dependence could be added to the model for greater realism, though it is not immediately clear how it should best be parametrized in the

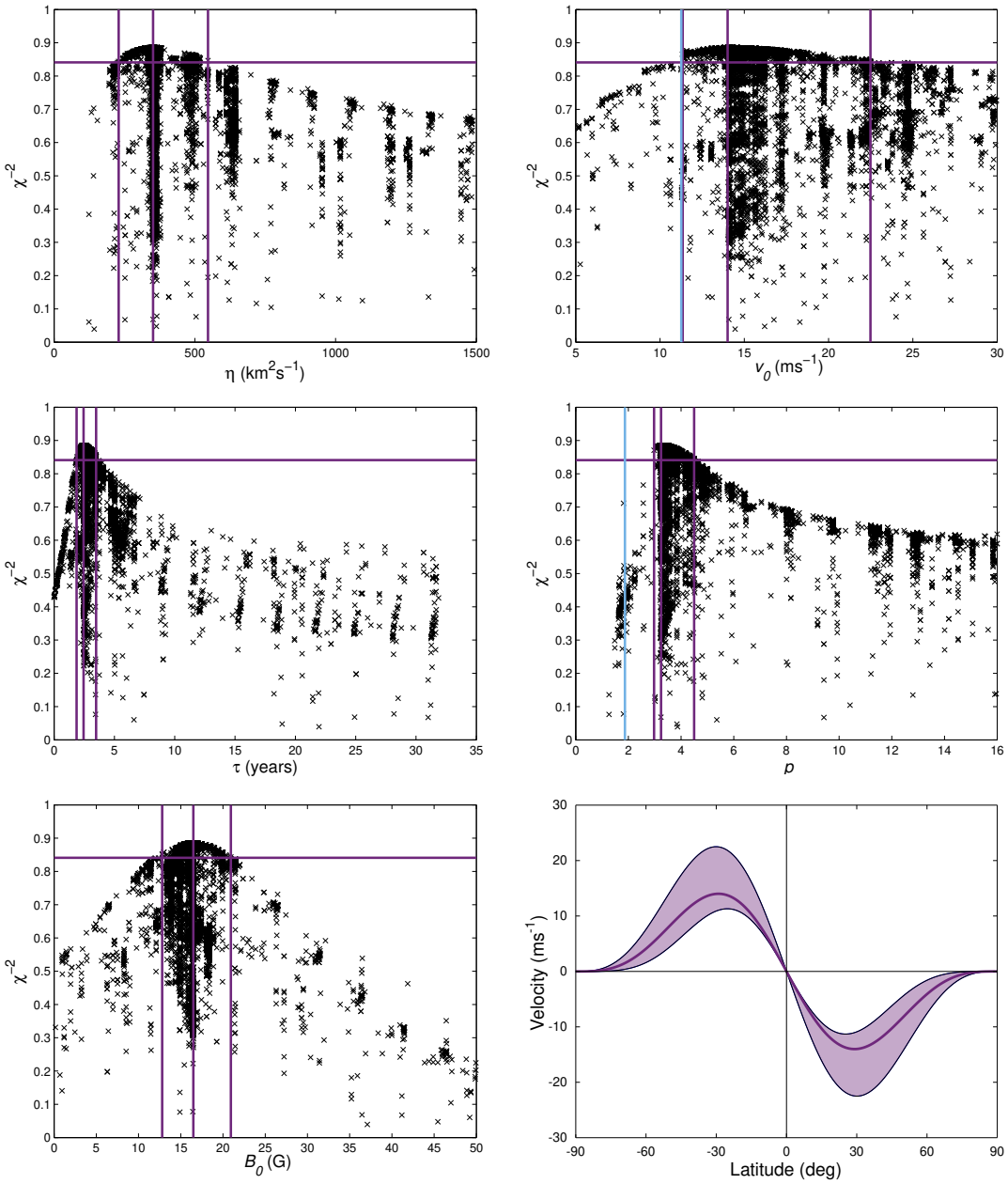

Figure 2.9: Scatter plots of every population member for each parameter. The horizontal purple line denotes 95% of the maximum χ^{-2}. The central vertical purple line is the optimum value for each parameter, with error bars given by the neighbouring vertical purple lines. The vertical blue lines in the top and middle right panels are the values obtained from fitting the velocity profile in Equation 2.1.24 to observational data (see Section 2.3). The bottom right panel shows the optimal meridional flow profile (bold purple) with acceptable profiles represented by the surrounding purple shading.

optimization. Furthermore, Hathaway & Rightmire (2010) noted that many SFT models use meridional flows which go to zero poleward of ±75° latitude which is not necessarily what is observed, as well as other deviations from observations. Upton & Hathaway (2014*a*) prescribed a profile with a maximum velocity of $12\,\mathrm{m\,s^{-1}}$ and Baumann *et al.* (2004) used $11\,\mathrm{m\,s^{-1}}$. Yeates (2014) found that a range of 11–15 $\mathrm{m\,s^{-1}}$ improves butterfly diagram correlation, and Wang *et al.* (1989*b*) and Wang & Sheeley (1991) found that a range of 7–13 $\mathrm{m\,s^{-1}}$ was acceptable. Wang *et al.* (2002*b*) found that a maximum velocity of 20–25 $\mathrm{m\,s^{-1}}$ accurately reproduced solar cycle features, although they used a profile which differs significantly from observations.

The parametric meridional flow profile in Equation 2.1.24 was also used by Muñoz-Jaramillo *et al.* (2009). They obtained a value of $p = 2$ by taking an average of helioseismic data weighted by density and fitting it to the sinusoidal profile. In this case $p = 2$ does not quite fall into the narrow acceptable range for p. The bottom right panel of Figure 2.9 shows that values within this range generally correspond to a peak velocity at ±30° before slowing down to $0\,\mathrm{m\,s^{-1}}$ at ±75°. As discussed in the above paragraph, this is not necessarily in line with observations.

Taking every member of the population above the $95\%\chi^{-2}_{\mathrm{max}}$ threshold, we find that the Pearson's correlation coefficient between the acceptable values for v_0 and p is $r = 0.86$, indicating that increasing the maximum velocity of the meridional profile generally requires an increase in p (see Figure 2.10). A faster velocity means that active regions are transported away from the equator quicker. To counteract this, a larger value of p narrows the band of latitudes at which the velocity is fast, and additionally brings the maximum velocity closer to the equator.

Another interesting result is that of 2.4 yr for the exponential decay time τ. Schrijver *et al.* (2002) found that a decay time of 5–10 yr was necessary to replicate regular polar field reversal, and Yeates (2014) found that a decay time of 10 yr produced a better fit between real and simulated butterfly diagrams. Lemerle *et al.* (2015) found that exponential decay did not have a large effect on the polar field reversal and decided to set $\tau = 32\,\mathrm{yr}$, effectively removing the decay term from the model.

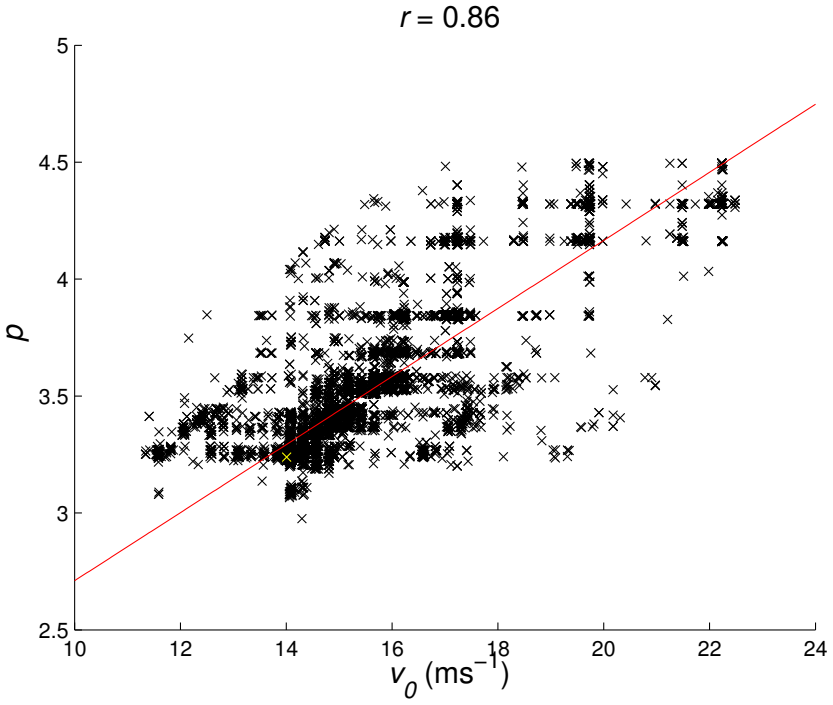

Figure 2.10: Scatterplot of acceptable values for v_0 against acceptable values for p from the 1D optimization in Table 2.5(a). The yellow cross highlights the optimal value and the red line is the linear regression line.

However, our optimal value for τ is close to the lower prescribed limit. This could be because of the model trying to account for the unusually weak polar field at the end of Cycle 23, while, for example, Lemerle *et al.* (2015) performed the optimization for Cycle 21. Wang *et al.* (2002*a*) overcame the problem of producing regular polar field reversals by increasing the meridional flow speed for stronger cycles.

Figure 2.11 highlights the need for the decay term in the 1D model when modelling Cycle 23 using this model. The axial dipole moment is defined as:

$$D\left(t\right) = \frac{3}{2}\int_0^{\pi} \overline{B}\left(\theta, t\right)\cos\theta\sin\theta\, d\theta, \tag{2.1.30}$$

and represents the net flux imbalance between the two hemispheres. We shall explore the dipole moment further in Chapter 4. The purple curve represents the axial dipole moment obtained using the optimal parameter set in Table 2.4(a), which provides a better fit to the observed axial dipole moment (blue) than the peach curve, which is produced from the same parameter set but with the decay term omitted. In this

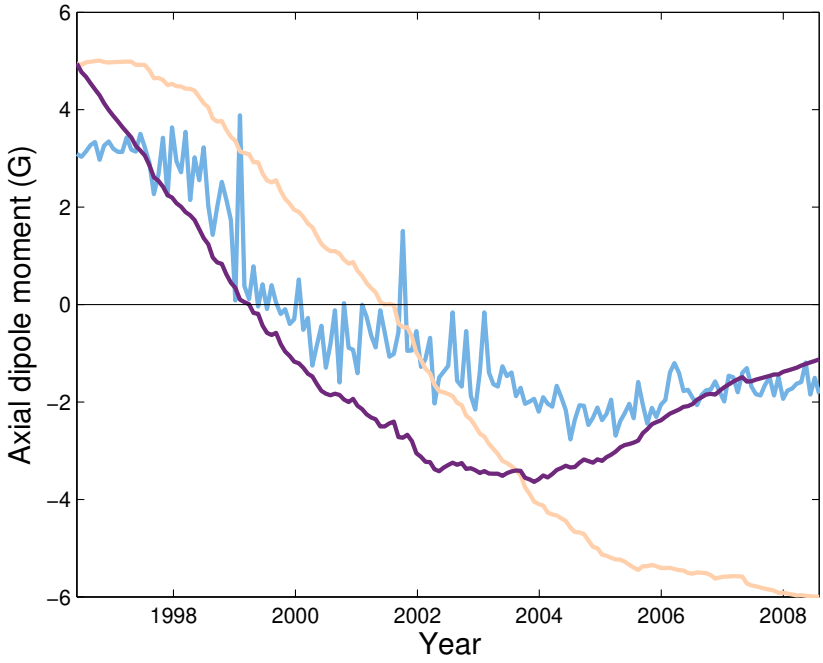

Figure 2.11: Axial dipole moments calculated from observed data (blue), the parameter set in Table 2.4(a) (purple), and the same parameter set but with the decay term omitted (peach).

case the polar field becomes too strong and is not weakened enough without the additional decay term. Jiang *et al.* (2015) found that the decay term was not required to obtain a close match between observed and simulated axial dipole moments, when using active region data from Li & Ulrich (2012). As well as using different active region data, a reduction in tilt angles and a smaller value of $\eta = 250\,\mathrm{km^2\,s^{-1}}$ were included to account for the lack of the decay term. If we use similar parameters for the 1D model, a better axial dipole moment fit is indeed obtained, but at the expense of an accurate butterfly diagram. Hence we stress that the optimal values in Table 2.4 are with respect to the measure of choice in Equation 2.1.25, and other choices of metric might give different results.

Of course, the choice of decay term is not independent of the other parameters, and the Pearson's correlation coefficient between the acceptable values of v_0 and τ is $r = 0.81$: an increase in the flow speed corresponds to less trailing flux being transported to the poles, so a fast decay to weaken the polar field would not be

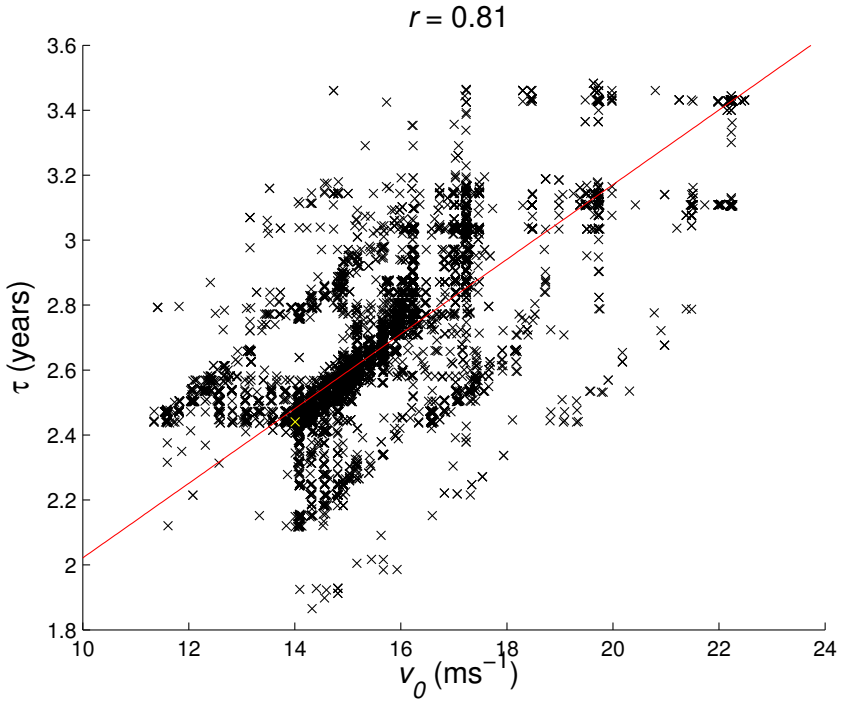

Figure 2.12: Scatterplot of acceptable values for v_0 against acceptable values for τ from the 1D optimization in Table 2.5(a). The yellow cross highlights the optimal value and the red line is the linear regression line.

required in the presence of a faster flow (Figure 2.12).

It should be noted that the decay term in Equation 2.1.1 is not directly observed. Baumann *et al.* (2006) gave a physical explanation of the decay term; namely, it is the effect of radial (i.e., inward) diffusion of flux into the solar interior, which cannot be accounted for directly in the SFT model. In spherical harmonics, different modes decay at different rates, whereas in the exponential decay used by Schrijver *et al.* (2002), all modes decay at the same rate. Baumann *et al.* (2006) found that the lowest-order mode decayed the slowest at a rate of 5 yr (with a corresponding volume diffusion of $\eta = 100\,\mathrm{km^2\,s^{-1}}$), in good agreement with the findings of Schrijver *et al.* (2002). When we include this more sophisticated form of radial diffusion in our model and perform the optimization, we find the lowest-order mode to have an optimal decay time of $\tau_1 = 2.7\,\mathrm{yr}$ (with a corresponding volume diffusion of $\eta = 190\,\mathrm{km^2\,s^{-1}}$), in good agreement with the decay time found in Table 2.5(a). Because of this good agreement, we opt to continue to use the original exponential

decay parameter. Radial diffusion is discussed in much more detail in Chapter 5.

The optimal value for B_0 is significantly higher than that used to approximate the initial profile in Figure 2.6. This might be attributed to the choice of functional form in Equation 2.1.22; not enough flux is prescribed between $\pm 45°$ and $\pm 80°$, so the algorithm compensates for this by increasing the maximum flux at $\pm 90°$. Alternatively, a strong initial polar field is also required to counteract the short decay time needed to reproduce the weak polar field at the end of Cycle 23.

2.1.9 Tilt Angles

Some studies (e.g. Jiang *et al.*, 2011; Yeates, 2014) found that multiplying the observed tilt angle of each BMR by a scaling factor reduces the polar field strength and improves polar field reversal, since the reduced tilt inhibits equatorial cross-cancellation and hence each magnetic region will contribute less to the axial dipole moment. To test this, a multiplicative tilt angle factor γ is included here as an extra parameter to be optimized within the range $0 \leq \gamma \leq 1.5$. Table 2.5(b) shows the results for the 6 parameter case, with the corresponding butterfly diagram in the top panel of Figure 2.13.

The optimal value of 0.55 for γ is lower than that found by Yeates (2014) ($\gamma \sim 0.8$) and Jiang *et al.* (2011) ($\gamma \sim 0.72$). It predictably produces a weaker polar field than in the case above where it wasn't included. Given that the main aim of the algorithm is to reduce differences between the real and simulated butterfly diagram pixels, it is reasonable to expect that the optimization algorithm will rely heavily on diffusion and high amplitudes of meridional flow to achieve weak polar fields, although it should be noted that this effect is reduced by the weighting in σ. Introducing the tilt angle factor as a means of reducing the polar field allows for the decay time to increase and the maximum meridional flow velocity to decrease, suggesting a delicate balance between the parameters and the roles they play in the model. While the polar field strength is better approximated in this case, the active regions are much

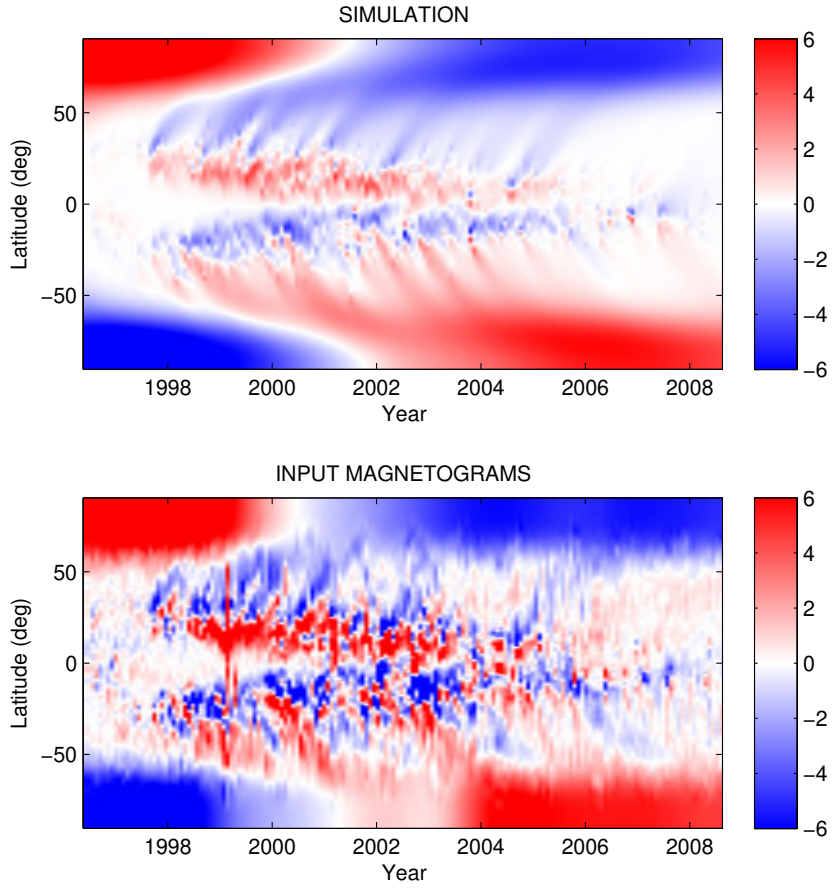

Figure 2.13: Top: Butterfly diagram for the optimal parameter 6-set for the 1D model with reduced tilt angles in Table 2.5(b). Bottom: Ground truth data for Cycle 23.

weaker than in the 5-parameter case, and polar field reversal occurs later in the simulation in both hemispheres. The fitness value of $\chi^{-2} = 1.09$ is above the 99% interval given in Section 2.1.7, which seems to indicate that the model matches the observations better than a randomly chosen map from the observed distribution. This is plainly a limitation of the χ^{-2} statistic; in particular, it likely indicates the presence of a significant σ_{model} term possessing a more complex structure over θ and t. In principle, it could be caused by too large a prescribed σ_{obs}, or by the relatively strong assumption of independence, or possibly over-fitting of the model, although the latter is unlikely given the small number of parameters in the model. It should be noted that the scaling of tilt angles is not a physical phenomenon, rather a method of reducing the flux in the model, though Cameron *et al.* (2010) argued that scaling the tilt angle by a factor of 0.7 mimics the effect of inflows around active regions. Moreover Dasi-Espuig *et al.* (2010) found an inverse correlation between cycle strength and tilt angle, suggesting that tilt angle variation plays a significant role in polar field variation.

2.2 Two-dimensional surface flux transport model

Yeates *et al.* (2015) developed a 2D model[e] which assimilates specific shapes of magnetic regions into the simulation on the day of emergence. The aim of the model is to better assimilate strong, multipolar regions, which are not accurately portrayed in a simpler bipolar form, as in the 1D model above, with the hope of simulating a more realistic photospheric field. This selection feature requires the model to be 2D. The model is fully automated, providing consistent highlighting of strong magnetic regions, and is designed to replace pre-existing regions rather than superimposing new ones.

[e]https://github.com/antyeates1983/sft_data

Each region is assimilated on the day when its centroid crosses the central meridian. The assimilation algorithm is described fully in the Appendix of Yeates *et al.* (2015). Briefly, the synoptic magnetograms are corrected for flux imbalance, then their absolute value is smoothed with a Gaussian filter (standard deviation $\sigma_{\text{smooth}} = 3$), so as to merge positive and negative polarities. Each region is then determined by a connected group of pixels above a threshold B_{par}. These pixels (from the original unsmoothed synoptic map) are then inserted into the simulation, replacing any pre-existing B_r in that region. The flux is corrected so as to preserve the pre-existing net flux in that region of the simulation.

Again we solve the SFT equation for the radial component of the magnetic field, as given in Equation 2.1.1. Rather than using a spectral method like the 1D model, we use $\mathbf{B} = \nabla \times \mathbf{A}$ to write

$$B \equiv B_r = \frac{1}{R_\odot \sin\theta} \left(\frac{\partial}{\partial \theta} \left(\sin\theta A_\phi \right) - \frac{\partial A_\theta}{\partial \phi} \right), \tag{2.2.1}$$

and solve the evolution equations for the vector potential:

$$\frac{\partial A_\theta}{\partial t} = \omega\left(\theta\right) R_\odot \sin\theta \, B - \frac{\eta}{R_\odot \sin\theta} \frac{\partial B}{\partial \phi} + S_\theta\left(\theta, \phi, t\right), \tag{2.2.2}$$

$$\frac{\partial A_\phi}{\partial t} = -v\left(\theta\right) B + \frac{\eta}{R_\odot} \frac{\partial B}{\partial \theta} + S_\phi\left(\theta, \phi, t\right). \tag{2.2.3}$$

These are solved in the Carrington frame (sidereal rotation period of 25.38 days) using a finite-difference method on a spatial grid of 180 cells equally spaced in sine-latitude and 180 cells equally spaced in longitude. We use an Euler timestepping method and periodic boundary conditions. While higher-order implicit methods would be more accurate, the Euler method is suitable for our purposes. Numerical problems might arise were we to assign a non-linear diffusion, but we will keep η constant in this model. Furthermore, it would be non-trivial to implement the replacement aspect of the emergence in an implicit scheme, as this is technically non-linear.

While differential rotation averaged out and so played no role in the axisymmetric

1D model, it must be included in the 2D case. Unlike meridional flow, differential rotation is well constrained by observations, and in the model we use the parametrization of Snodgrass & Ulrich (1990) for the angular velocity in the Carrington frame:

$$\omega(\theta) = 0.521 - 2.396\cos^2\theta - 1.787\cos^4\theta \,\mathrm{deg\,day^{-1}}. \tag{2.2.4}$$

For the parameter B_{par}, and for GONG input data, Yeates *et al.* (2015) chose the threshold of $B_{\mathrm{par}} = 15\,\mathrm{G}$ in order that the difference between the observed unsigned flux and simulated unsigned flux (due to the smoother magnetic field distribution) remained approximately constant. Here, this parameter is added to the optimization. If given enough freedom, the algorithm would gradually reduce B_{par}, allowing more and more magnetic regions to be inserted until the original synoptic map is essentially copied in (analogous to $B_{\mathrm{par}} \sim 0\,\mathrm{G}$). To avoid this, the lower bound is set at 10 G with an upper bound of 50 G. Figure 2.14 shows snapshots of 1928 CR from four simulations with alternative values of B_{par} between 10 G and 50 G, and all other parameters fixed. As the threshold B_{par} increases, fewer active regions are assimilated into the simulation.

2.2.1 Five-parameter optimization

The synoptic magnetograms from NSO Kitt Peak are used to identify strong regions for assimilation. For simplicity, Yeates *et al.* (2015) did not incorporate exponential decay into the model as in Equation 2.1.1. We perform optimization runs for the model both without decay and with the decay term included. Initially we consider the former case. Aside from B_{par}, parameters are given the same upper and lower limits as in Section 2.1.6. Table 2.5(c) shows the results of the optimization. The corresponding butterfly diagram is shown in the middle panel of Figure 2.15, with the optimal 1D butterfly diagram from Section 2.1.8 in the top panel for direct comparison.

The 2D model qualitatively improves the butterfly diagram, with active regions

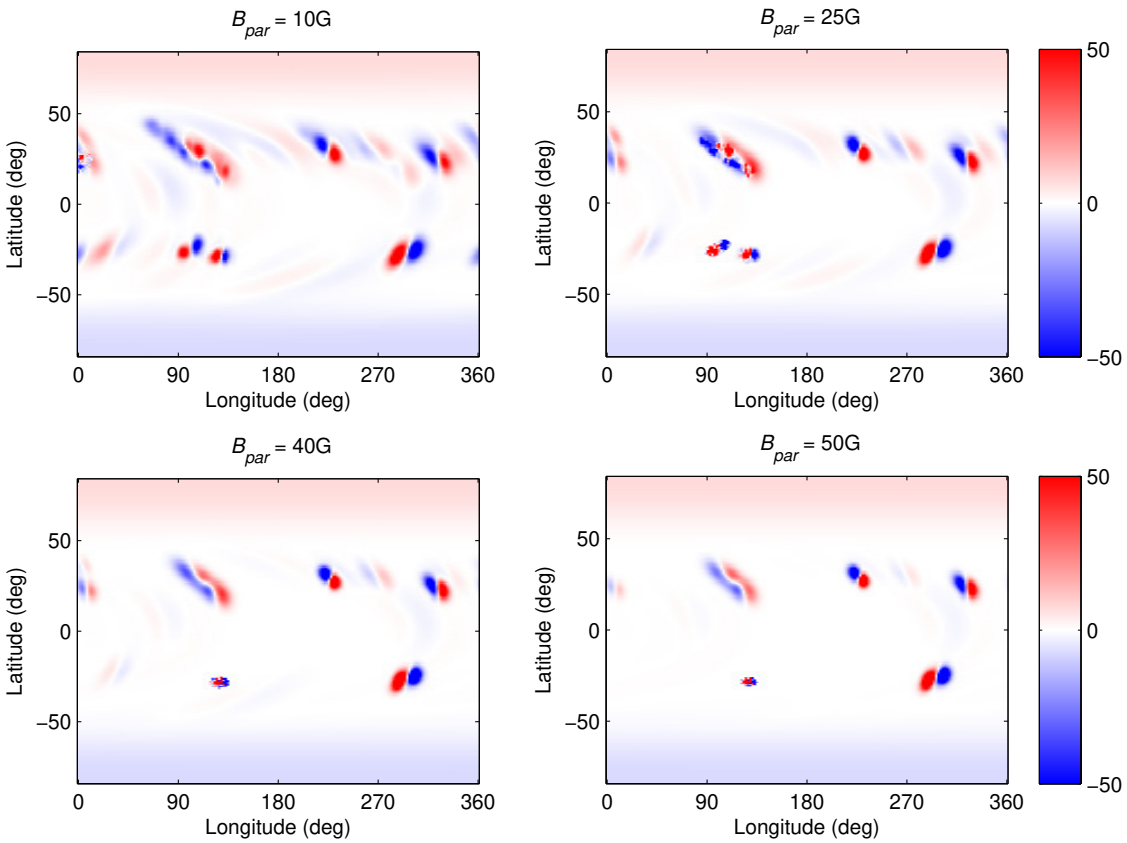

Figure 2.14: Four snapshots of 1928 CR from simulations with active regions selected by different values of the magnetic flux density threshold $B_{\rm par}$ and all other parameters fixed. Here $B_{\rm par}$ increases from left to right and top to bottom.

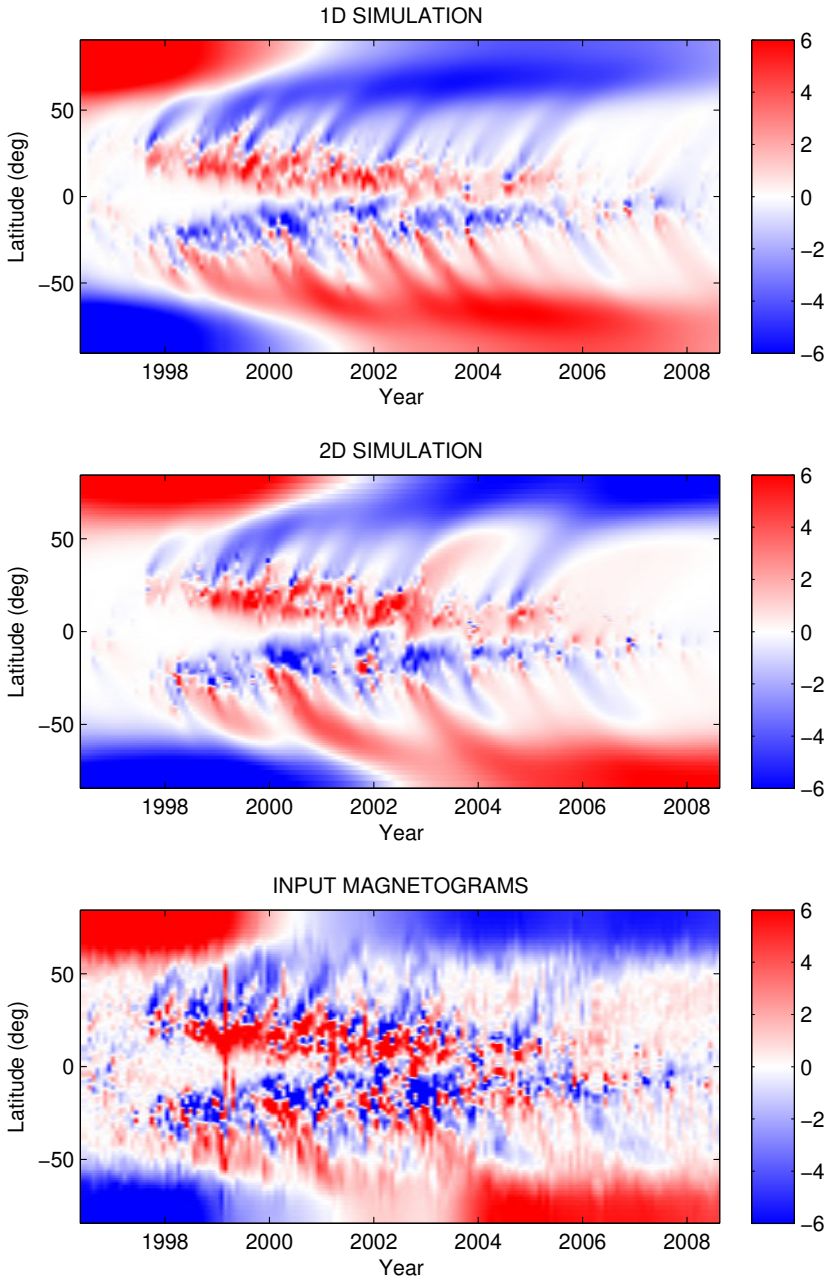

Figure 2.15: Top: Butterfly diagram for the optimal parameter 5-set for the 1D model in Table 2.5(a). Middle: Butterfly diagram for the optimal parameter 5-set for the 2D model with varying $B_{\rm par}$ in Table 2.5(c). Bottom: Ground truth data for Cycle 23.

predictably more accurate, leading to the inclusion of more poleward surges in the simulation which can be identified in the observed butterfly diagram (though the gradient and strength of each surge is not always correct), and a more realistic polar field. The optimal parameters in Table 2.5(c) are within the range of other results from simulations and observations described in Section 2.1.8. A diffusivity of $\eta = 455.6\,\mathrm{km^2\,s^{-1}}$ is stronger than in the 1D model, but the inclusion of an exponential decay term is expected to reduce this. An increased diffusivity is somewhat supported by Virtanen *et al.* (2017), who used a value of $\eta = 400\,\mathrm{km^2\,s^{-1}}$ in the same 2D model but for a single simulation of multiple cycles. The range and optimal value for v_0 is lower than for the original 1D case, indicating that there can be inherent differences between models. Moreover, Virtanen *et al.* (2017) found that a value of $v_0 = 11\,\mathrm{m\,s^{-1}}$ correctly reproduced shapes of poleward surges and polar fields, in excellent agreement with our optimal value.

Figure 2.16 shows every generated value of B_{par} against χ^{-2}. The central vertical line indicates the optimum value of 39.8 G, with the left and right vertical lines denoting the acceptable range for B_{par}, as in Figure 2.9. For the remainder of the 2D optimizations, B_{par} is fixed at the optimal value of 39.8 G to attain consistency, unless stated otherwise. This should ensure that only newly emerging regions are inserted for each Carrington rotation. However, the presence of the strong mid-latitudinal region of positive flux in the northern hemisphere around 2003–2004 could be attributed to the choice of large B_{par}, since smaller regions of negative flux which would otherwise cancel out this positive flux are not being assimilated. The bottom left panel of Figure 2.14 closely represents the scenario when B_{par} is set at its optimal value. Virtanen *et al.* (2017) used a threshold of $B_{\mathrm{par}} = 50\,\mathrm{G}$, and this lies just outside of our acceptable range. Comparing the bottom two panels of Figure 2.14, however, shows that the differences between our optimal value and their chosen value are minor.

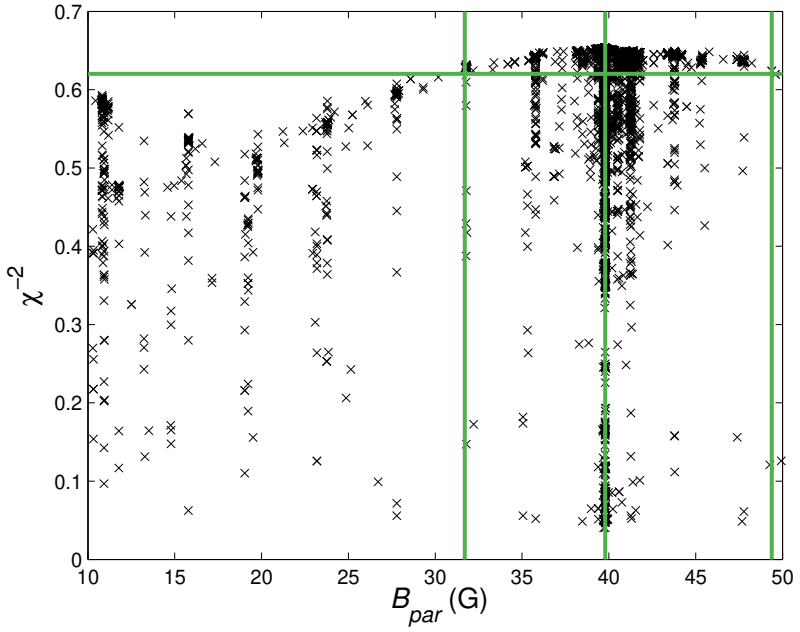

Figure 2.16: Each population member for the 5-parameter optimization of the 2D model in Table 2.5(c), with $B_{\rm par}$ restricted to $B_{\rm par} \geq 10\,{\rm G}$. The horizontal green line denotes 95% of the maximum χ^{-2}. The central vertical green line is the optimum value for each parameter, with error bars given by the neighbouring vertical green lines.

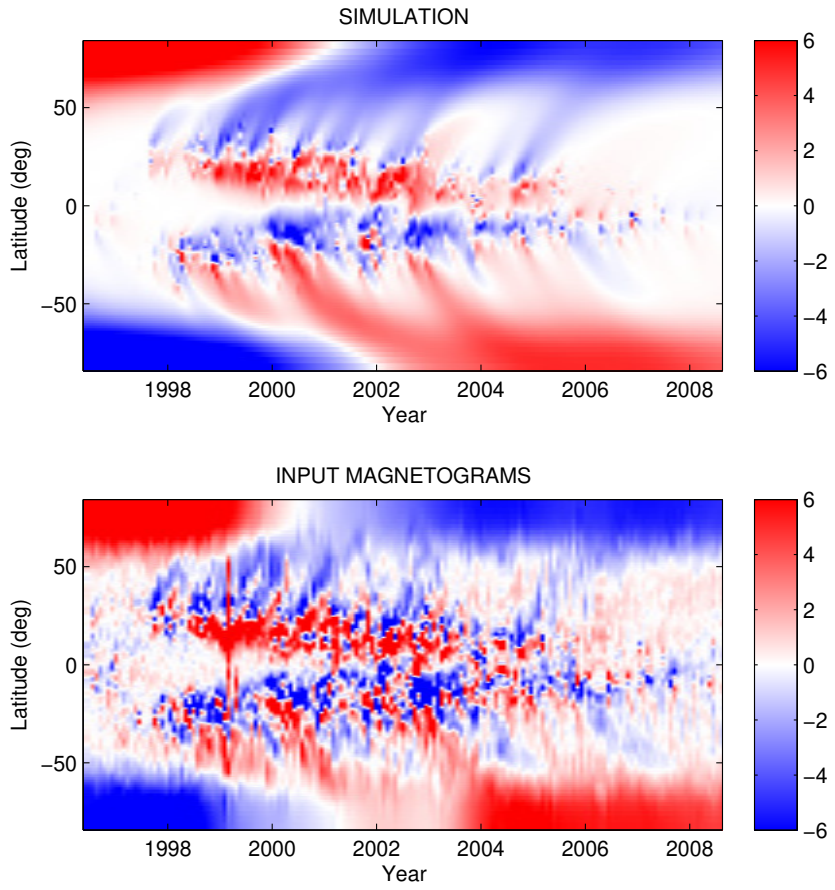

Figure 2.17: Top: Butterfly diagram for the optimal parameter 5-set for the 2D model with varying τ in Table 2.5(d). Bottom: Ground truth data for Cycle 23.

2.2.2 Incorporating exponential decay

As discussed above, the decay parameter τ was originally added to the SFT model to produce regular polar field reversals. The 2D model did not initially take account of this exponential decay, but we incorporate it to assess whether the optimal value in Table 2.5(a) is reasonable.

As shown in Figure 2.17, including the decay term improves timing of polar field reversal by a couple of months, but is not enough to replicate precisely the observed reversal time. Poleward surges are generally wider in the simulation, leading to the reduction of some mid-latitude features, most notably the strong surge of positive flux around 2003–2004 in the northern hemisphere, which is more visible in Figure 2.15.

The optimization results are shown in Table 2.5(d). Surprisingly, the addition of an extra decay term induces a minimal decline in diffusion, and it is not enough to bring it down to $351.6\,\mathrm{km^2\,s^{-1}}$ as found in the 1D case. Rather, B_0 increases to account for the stronger decay of the polar fields in this regime. Most significantly, we obtain an optimal value of $\tau = 4.5\,\mathrm{yr}$. This is higher than the optimum found in the 1D model and in closer agreement with Schrijver *et al.* (2002), although the acceptable range is considerably wider towards the upper limit, indicating that a decay term may not be required in the assimilative model. This is supported by the value of χ^{-2} which does not increase significantly with the addition of the decay term. Furthermore, Figure 2.18 shows that the axial dipole moments calculated using the optimal parameter sets for the 2D model, with and without the exponential decay term (brown and green curves respectively), both produce good fits to the observed profile (blue). This indicates that the method of new flux assimilation in the 2D model is better able to account for the weak polar field at the Cycle 23/24 minimum than the idealized BMRs used in the 1D model, since it does not require an additional decay term. Coupled to the short optimal decay timescale are smaller optimal values for v_0 and p, suggesting that the relationships and correlations discussed in Section 2.1.8 also hold for the 2D case.

2.3 Comparison with meridional flow observations

Although observations of the meridional flow are not yet fully reliable, we can use the data that are available to give an independent test of the optimized model.

David Hathaway kindly provided us with measurements of the meridional flow for Solar Cycle 23, calculated by tracking features in images from MDI. The data were supplied as coefficients of the following parametrization:

$$v\left(\theta\right) = \left(C_0 + C_1\cos\theta + C_2\cos^2\theta + C_3\cos^3\theta + C_4\cos^4\theta + C_5\cos^5\theta\right)\sin\theta. \quad (2.3.1)$$

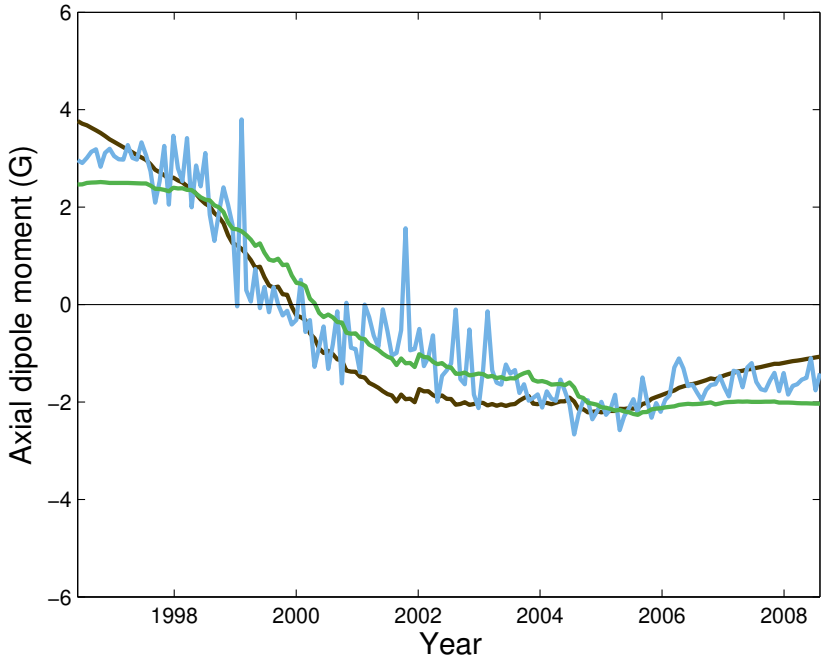

Figure 2.18: Axial dipole moments calculated from observed data (blue), the parameter set in Table 2.4(c) (green), and the parameter set in Table 2.4(d) (brown).

The meridional flow measurements for each Carrington rotation are shown in Figure 2.19 (blue curves). The observations tend to follow either a fast or slow flow, highlighted by denser blue areas, indicating the dependence on time and that the flow transitions between the two extremes throughout the cycle. Additionally, for a small number of Carrington rotations an equatorward counterflow is observed at high latitudes, though it should be noted that such a counterflow was not visible in HMI data (Hathaway & Upton, 2014). The choice of flexible profile in Equation 2.1.24 does not allow for this phenomenon.

The optimal profile using the parameters from the 1D optimization in Table 2.5(a) is shown in purple in Figure 2.19 for comparison. Whilst the observed and optimal profiles are similar in shape, the optimal profile is too fast and reaches its peak at a slightly lower latitude. Moreover, the observed profiles tend to extend beyond $\pm 75^\circ$ but the optimal profile chooses to go to zero throughout the polar regions, giving a possible explanation as to why many SFT models incorporate this feature. Furthermore, the 1D optimal profile remains almost completely within the bounds

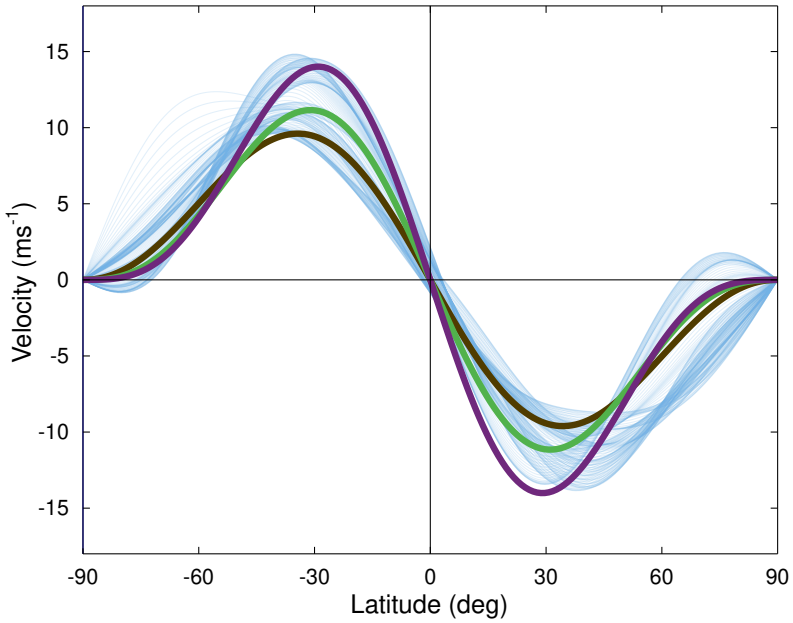

Figure 2.19: Comparison of various meridional flow profiles: observed for each CR (blue), 1D optimum (purple), 2D optimum (green) and 2D optimum with decay (brown).

given by the observations, excluding at its peak in the northern hemisphere for which asymmetry in the observations can be held responsible.

The green and brown profiles in Figure 2.19 represent the optima for the 2D model excluding and including exponential decay respectively. Both profiles are fully contained within the observational limits, except for a small section of the brown curve in the southern hemisphere which is due to a lower than average maximum velocity. Of the three optimal profiles, the 2D regime without decay matches the average observed profile the closest, whilst the decay-enhanced flow is slightly slower (though Hathaway & Rightmire (2010) observed speeds of $8\,\mathrm{m\,s^{-1}}$ at cycle maximum). It does, however, continue to latitudes poleward of $\pm 70°$, almost emulating the observational data. One limitation of tracking magnetic features to measure the meridional flow is that it is not always easy to distinguish between the effects of the meridional flow and the effects of supergranular diffusion. For this reason, flows derived from feature tracking tend to peak at higher latitudes (e.g. Dikpati *et al.*, 2010, Figure 1), giving a possible explanation as to why the observed curves in Figure 2.19 tend

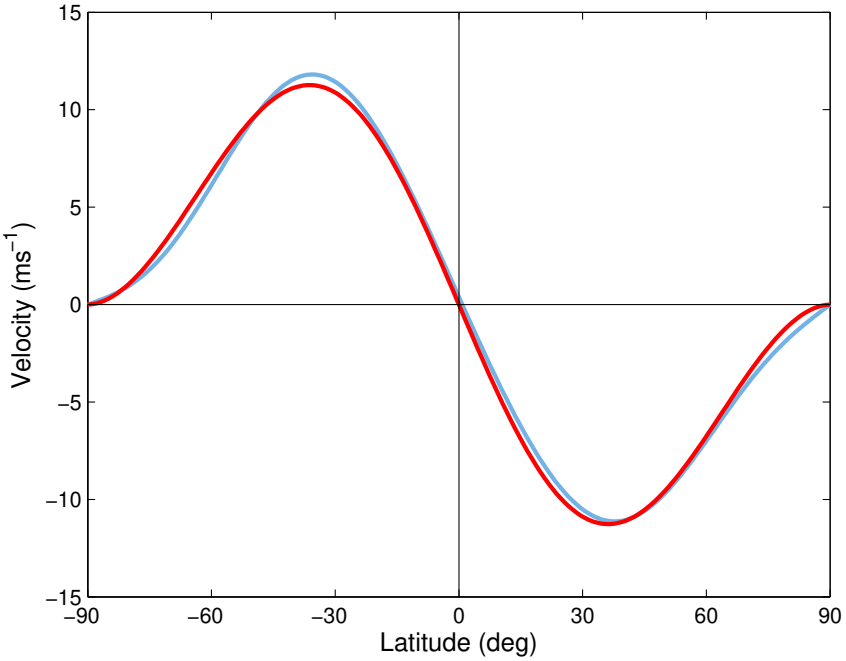

Figure 2.20: Comparison of average observed (blue) and fitted (red) meridional flow profiles.

to peak at higher latitudes than the modelled curves.

We use a non-linear least-squares fitting method to fit the parametrized form of the meridional flow in Equation 2.1.24 to the average observed coefficients given by David Hathaway to ensure it is actually possible to match the observed profile. The average observed and fitted profiles, shown in Figure 2.20 (blue and red respectively), match closely for $v_0 = 11.3\,\mathrm{m\,s^{-1}}$ and $p = 1.87$, and slight asymmetry in the average observed profile is confirmed. This value of p is close to that of Muñoz-Jaramillo *et al.* (2009) and is within the acceptable ranges for p in the above 2D regimes, but is outside the equivalent range in the 1D optimization run, whence we infer that the 1D model requires the maximum velocity to be closer to the equator than is observed.

Given that the parametrization is able to closely fit the observed data, we could fix one of the velocity-related parameters, say p, to the observed value and perform optimization runs for the two models. We choose p because the model is generally less sensitive to the choice of v_0, and $p = 1.87$ is outside the acceptable range for the

1D model.

The optimization results with p fixed in the 1D model are shown in Table 2.5(e). The value $p = 1.87$ corresponds to a maximum velocity at $\pm 35°$, meaning poleward transport is slower at low latitudes. This results in more flux cancellation across the equator and so more trailing flux is present in the transport regions, as observed in the top panel of Figure 2.21. This feature appears to be a common occurrence in the standard SFT model (cf. Figures 2.8 and 2.13). The upshot of this numerically is that the selected decay time of 1.9 yr is even shorter than in the original 1D case to counteract the large amounts of flux accumulating at the poles. This couples with a slow velocity, made even slower by the small value of p, adhering to the relationship found in Section 2.1.8. The timing of polar field reversal, meanwhile, is reproduced reasonably accurately. Except for a marginally smaller value of χ^{-2}, fixing p does not significantly hinder the quantitative performance of the 1D model, even though $p = 1.87$ is not in the acceptable parameter range for regime (a).

With the higher-latitudinal velocity peak and the absence of τ in the 2D model, the resulting diffusion value given in Table 2.5(f) is slightly larger than in previous regimes. Contrary to expectation, the optimal maximum velocity is higher than the previous 2D cases, but still with wide error bounds. Given that $p = 1.87$ lies within the acceptable range in regime (c), it is reasonable to expect that optimal values and associated ranges would be in line with results in Section 2.2 and hence observations and previous studies. Consequently the optimal butterfly diagram (middle panel of Figure 2.21) confirms this, offering only subtle changes to Figure 2.15, for example a polar field restricted to higher latitudes due to the increase in diffusivity.

2.3.1 Initial profiles

The choice of initial profile in Equation 2.1.22 was made to ensure consistency across all regimes. However, fixing the observed initial profile could in theory aid the algorithm in achieving the correct polar field reversal timing for each hemisphere

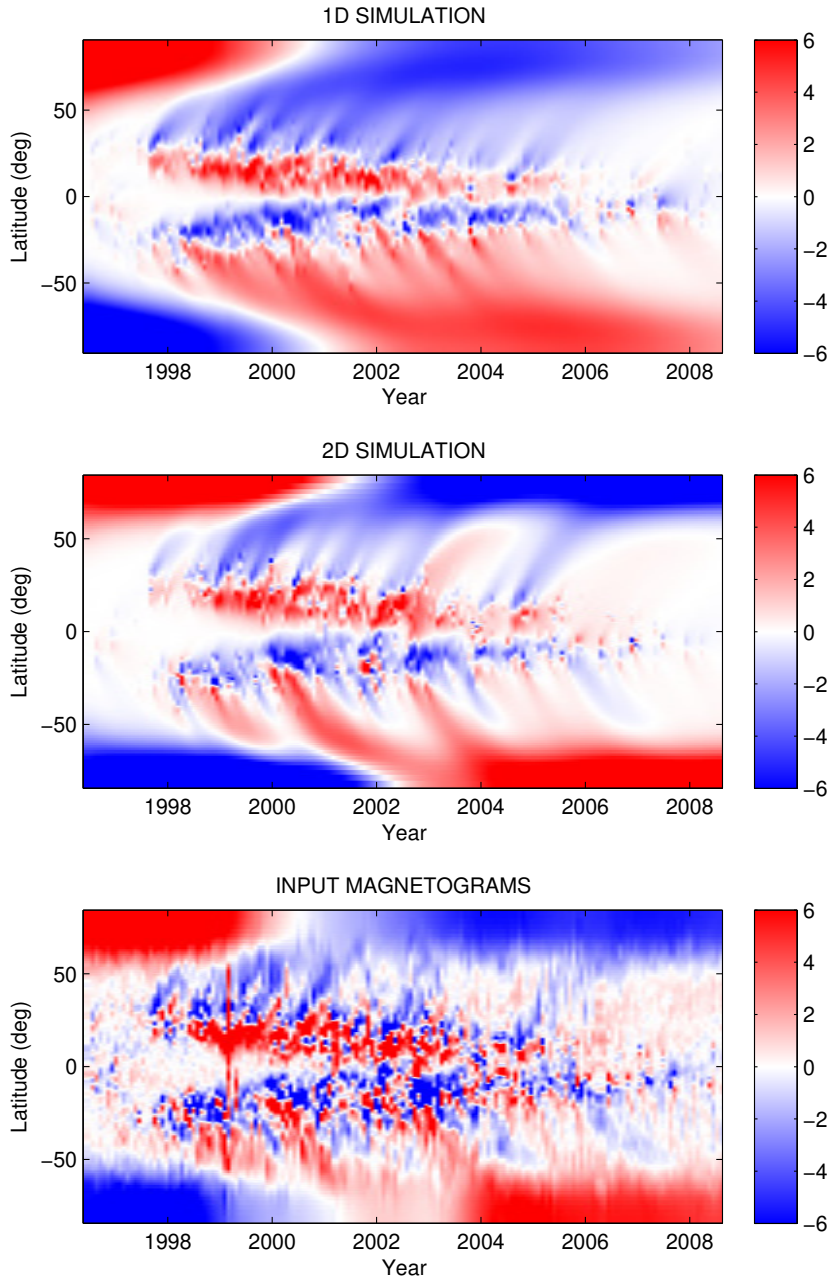

Figure 2.21: Top: Butterfly diagram for the optimal parameter 4-set for the 1D model with fixed $p = 1.87$ in Table 2.5(e). Middle: Butterfly diagram for the optimal parameter 3-set for the 2D model with fixed $p = 1.87$ in Table 2.5(f). Bottom: Ground truth data for Cycle 23.

and ensuring the initial field strength is more accurate, subject to observational errors.

Therefore we perform optimization runs for both models for Cycle 23 with the initial profile fixed to that observed (shown in blue in Figure 2.6). The optimal values and associated acceptable ranges can be found in Table 2.5(h) and (i) for the 1D and 2D models respectively. We find that imposing the observed profile does not improve the fitness value in either case. This might be expected since the parameter B_0 is removed from the optimization, meaning the procedure is more constrained. However, only the diffusion η and maximum meridional velocity v_0 are significantly affected by the change in initial profile, and even in these cases there is only a relatively small difference, meaning each optimal parameter lies comfortably within the acceptable ranges given in Table 2.5(a) and (c). Furthermore, there are no significant qualitative differences in the optimal butterfly diagrams (Figures 2.22 and 2.23). Figure 2.24 shows the percentage difference between each case for each model and we find that the main regions of discrepancy are around the initial advection of low-latitude regions (as would be expected) and the timing of the subsequent polar field reversal. The polar field at the end of the cycle and the majority of the rest of the diagram is largely unaffected by the change of initial profile. From this we conclude that the sinusoidal profile in Equation 2.1.22 is sufficient for this study.

2.4 Other solar cycles

With its automated assimilation of active region data, the 2D model can easily be adapted for other cycles, provided there are sufficient data available. Evaluations of Cycles 21, 22, and 24 (up to the end of 2015) using NSO data have been carried out to search for cycle-to-cycle variation.

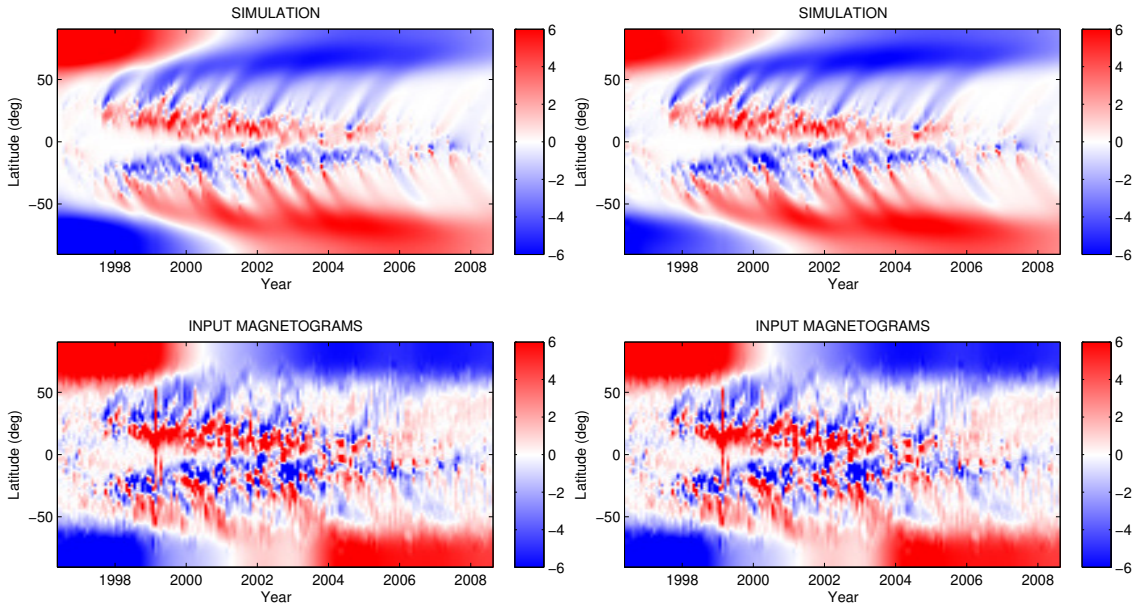

Figure 2.22: Comparison between optimal butterfly diagrams for the 1D model using the idealized initial profile (top left) and observed initial profile (top right).

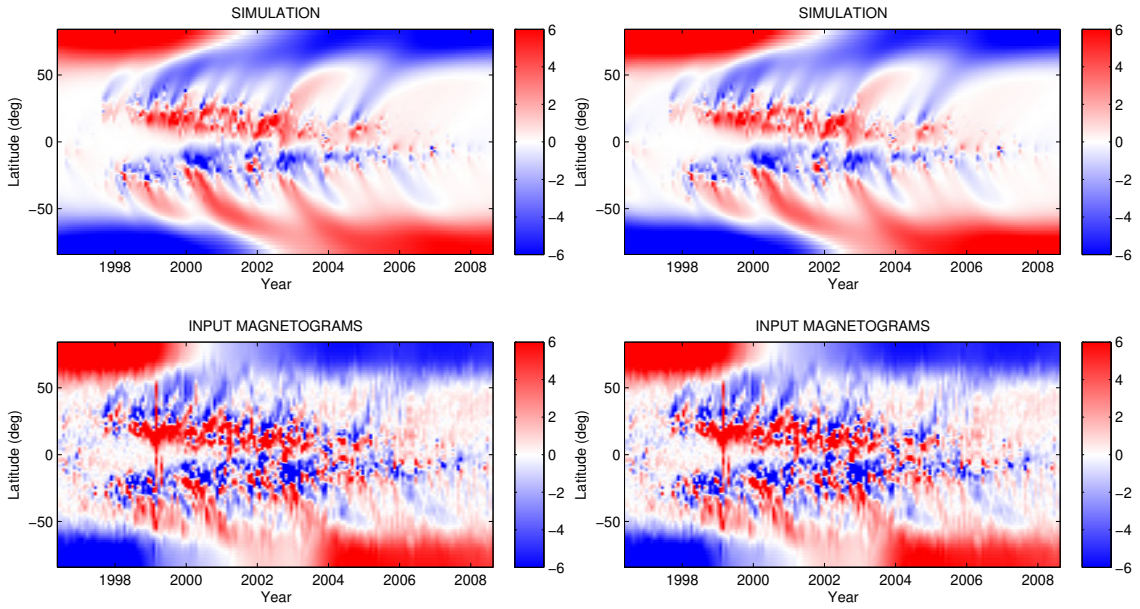

Figure 2.23: Comparison between optimal butterfly diagrams for the 2D model using the idealized initial profile (top left) and observed initial profile (top right).

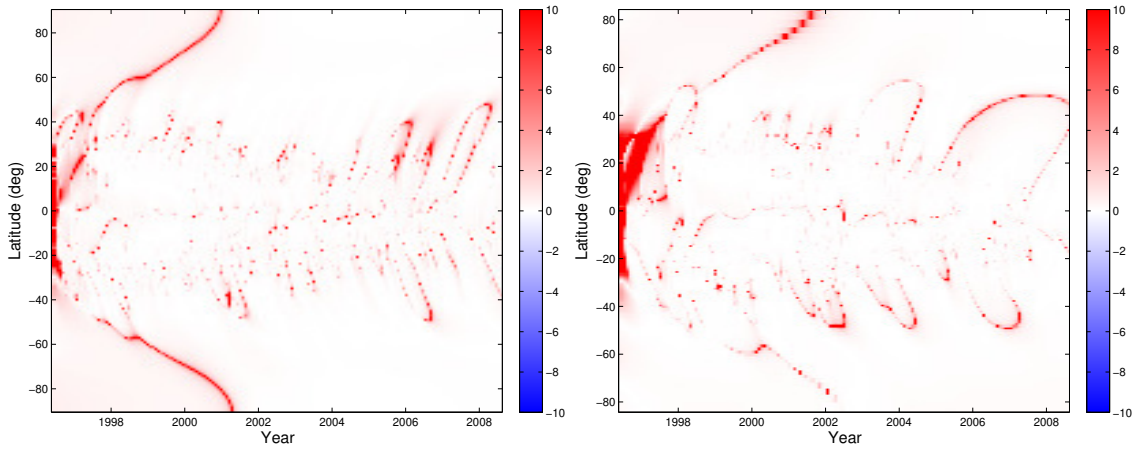

Figure 2.24: Percentage difference between the optimal butterfly diagrams using the two different initial profiles for the 1D model (left) and 2D model (right).

2.4.1 Cycle 21

Table 2.5(j) shows the optimum parameters for Cycle 21 (1976 May 1–1986 March 10). Both η and v_0 are in agreement with previous studies. Most notably, $v_0 = 9.2\,\mathrm{m\,s^{-1}}$ is slower than the maximum speed of Cycle 23, supporting Upton & Hathaway (2014*a*): a faster flow in Cycle 23 would have resulted in a weaker polar field at cycle minimum since leading flux would be taken away from the equator quickly and so would have less time to cancel across the equator. This optimum value, however, is just outside the range of 10–13.2 $\mathrm{m\,s^{-1}}$ as found by Komm *et al.* (1993) using feature tracking during Cycle 21. However, this range overlaps with a large portion of the 95% confidence interval obtained by the optimization population.

The interpolated NSO data are shown in the bottom panel of Figure 2.25 with the corresponding simulated butterfly diagram in the top panel of Figure 2.25. Aside from a negative-polarity observational artefact in the northern hemisphere halfway through 1979, many features of active regions are well reproduced. There are three instances of large concentrations of opposite flux being transported polewards in the northern hemisphere; the latter of these is over-estimated by the simulation and this could be attributed to the model incorrectly reading in the corresponding emergence region. Polar field reversal for both poles is too late in the model, particularly in

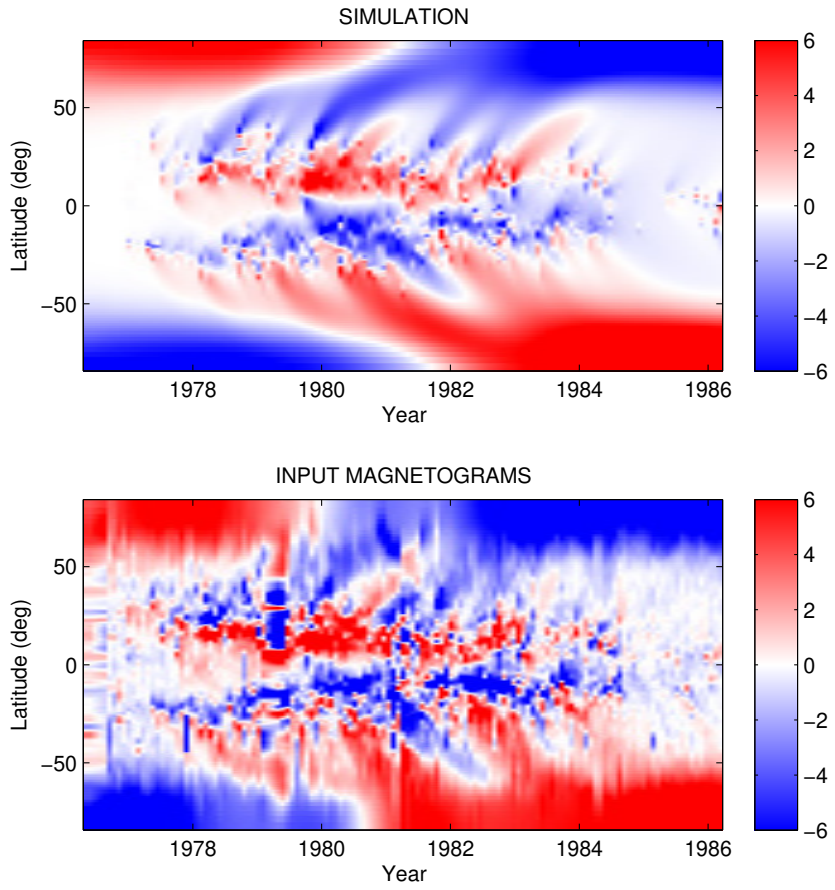

Figure 2.25: Top: Butterfly diagram for the optimal parameter 4-set for the 2D model in Table 2.5(j). Bottom: Ground truth data for Cycle 21.

the northern hemisphere where the difference is in the region of a year.

Lemerle *et al.* (2015) performed a similar optimization process for Cycle 21 using a 2D model and a BMR database compiled by Wang & Sheeley (1989). Although they used a different parametrization for the meridional flow and different sources of flux, their optimal parameter ranges for η and v_0 are in good agreement with those in Table 2.5(j). Our diffusion coefficient $\eta = 455.7\,\mathrm{km^2\,s^{-1}}$ lies within their acceptable range of 240–660 $\mathrm{km^2\,s^{-1}}$ and $v_0 = 9.2\,\mathrm{m\,s^{-1}}$ falls between 8–18 $\mathrm{m\,s^{-1}}$ as calculated by `PIKAIA` in their study. They used the following functional form to represent meridional flow:

$$v\left(\theta\right) = -v_0\,\mathrm{erf}^q\left(v\sin\theta\right)\mathrm{erf}\left(w\cos\theta\right). \tag{2.4.1}$$

Their optimization returned values of $v_0 = 12\,\mathrm{m\,s^{-1}}$, $q = 7$, $v = 2$ and $w = 8$. This

gave a profile similar to that of Wang *et al.* (2002*b*), but with a less extreme gradient at the equator. However, when normalized, the profile shape was comparable to the observed profile formed from Doppler measurements obtained by Ulrich (2010), and the observed profile lay well within the error bars for the optimal solution, except for some return flows at high latitudes, which were not incorporable in Equation 2.4.1, mirroring the limitation of our parametrization in Equation 2.1.24. Using a non-linear least-squares fitting method, we are able to attempt to fit the functional form in Equation 2.1.24 to the optimal meridional flow profile of Lemerle *et al.* (2015). The best fit corresponds to values of $v_0 = 13.6\,\mathrm{m\,s^{-1}}$ and $p = 3.88$. This value for v_0 is in agreement with observations and acceptable ranges for other regimes, but is above the range for Cycle 21. Despite lying within the acceptable range, $p = 3.88$ favours the high values for p obtained from optimization runs as opposed to the lower values extracted from observational data. This could suggest an inherent flaw within the SFT model whereby the model performs better when the maximum velocity is prescribed to be closer to the equator.

2.4.2 Cycle 22

Table 2.5(l) shows the optimization results for Cycle 22 (1986 March 10–1996 June 1). The fit is marginally worse than for Cycle 21, but optimal values for η and v_0 remain within in plausible ranges. The optimal diffusion in this case increases to $506.2\,\mathrm{km^2\,s^{-1}}$, but is in better agreement with Wang *et al.* (2002*b*). The optimal maximum velocity for Cycle 22 is even smaller than that of Cycle 21, further supporting the fact that a slower meridional flow results in a stronger polar field at cycle minimum, and explaining the high optimal maximum velocity for Cycle 23. van Ballegooijen *et al.* (1998) performed SFT simulations for Cycle 22 with $\eta = 450\,\mathrm{km^2\,s^{-1}}$ and $v_0 = 11\,\mathrm{m\,s^{-1}}$ which produced polar field strength in agreement with observations. Again, these values are in accordance with ranges given in Table 2.5(l).

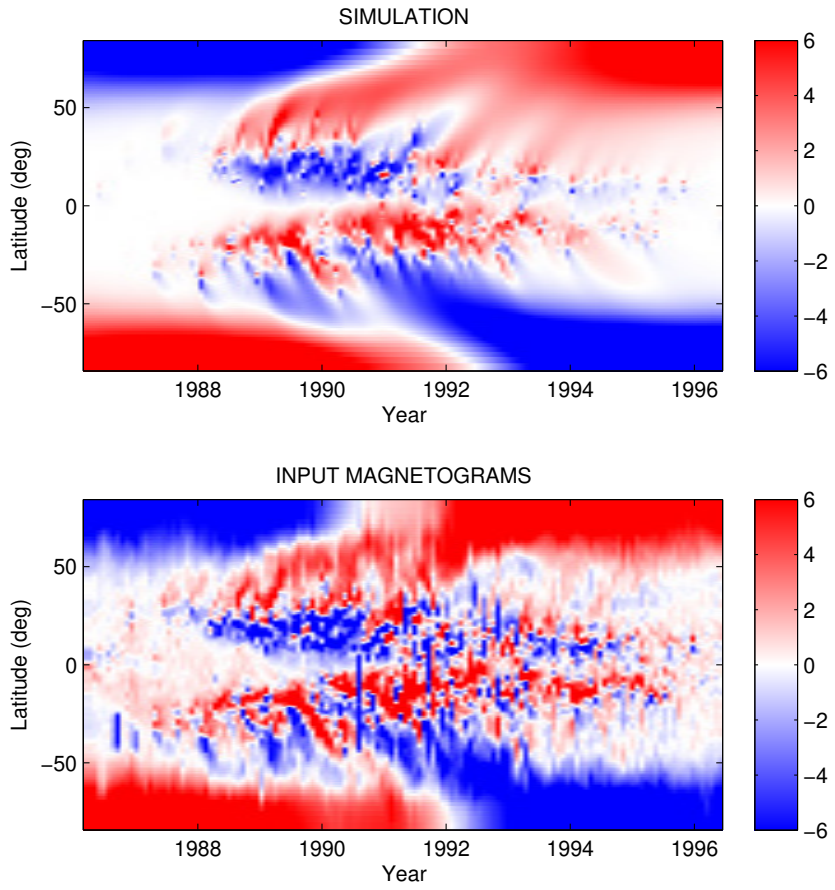

Figure 2.26: Top: Butterfly diagram for the optimal parameter 4-set for the 2D model in Table 2.5(l). Bottom: Ground truth data for Cycle 22.

The ground truth data are shown in the bottom panel of Figure 2.26 and the simulated butterfly diagram is in the top panel of Figure 2.26. The model has recreated polarity reversal much more successfully here, with only a slight delay in the north. Towards the end of the cycle there is a large build-up of positive flux and some weak, but visible, poleward surges in the northern hemisphere that have appeared in the simulation but are not clearly seen in the real butterfly diagram.

2.4.3 Cycle 24 (so far...)

Table 2.5(n) shows the results for the first half of Cycle 24 (2008 August 3–2016 Jan 1). We obtain a much higher value of χ^{-2} for Cycle 24 compared to previous cycles, but we suspect that this is might change once we have data for the full

cycle and try to replicate the long-term effects. The diffusivity $\eta = 454.6\,\mathrm{km^2\,s^{-1}}$ is within viable ranges found in literature, though the maximum velocity is close to the lower prescribed bound. The initial polar field $B_0 = 4.2\,\mathrm{G}$ is lower than in previous cycles as the model needs to replicate the weak polar field at the Cycle 23/24 minimum. Acceptable ranges of parameters are generally broad, but performing the optimization on the full cycle in the next few years should tighten the upper and lower bounds. Indeed, when a similar optimization process is performed on half of Cycle 23, the acceptable ranges are found to be wider, though the shorter time period has a negligible effect on the specific optimal values (see Table 2.5(g) for optimal values and ranges).

The interpolated Kitt Peak data are shown in the bottom panel of Figure 2.27 with the corresponding simulated butterfly diagram in the top panel of Figure 2.27. Although a large portion of the cycle is yet to take place, there are still some notable features, such as the prominent leading-polarity region in the northern hemisphere in 2011. This region was the primary subject of Yeates *et al.* (2015). Polar field reversal is slightly late in the simulated butterfly diagram; performing an optimization once the full cycle has completed might remedy this, though a region of negative polarity in the northern hemisphere at the end of the current dataset may not correctly be reproduced, unless the data are corrected.

2.4.4 Exponential decay in other solar cycles

Including exponential decay in the model for Cycles 21, 22, and 24 produces optimal values of $\tau = 10.2 \in [3.1, 32.0]\,\mathrm{yr}$, $\tau = 7.6 \in [3.1, 32.0]\,\mathrm{yr}$ and $\tau = 15.1 \in [2.5, 32.0]\,\mathrm{yr}$ respectively (Table 2.5(k), (m) and (o) respectively). These are in better agreement with Schrijver *et al.* (2002) and Lemerle *et al.* (2015), indicating that the low optimal value for τ may be necessary only for Cycle 23 in order to successfully reconstruct the unusually weak polar fields at Cycle 23/24 minimum.

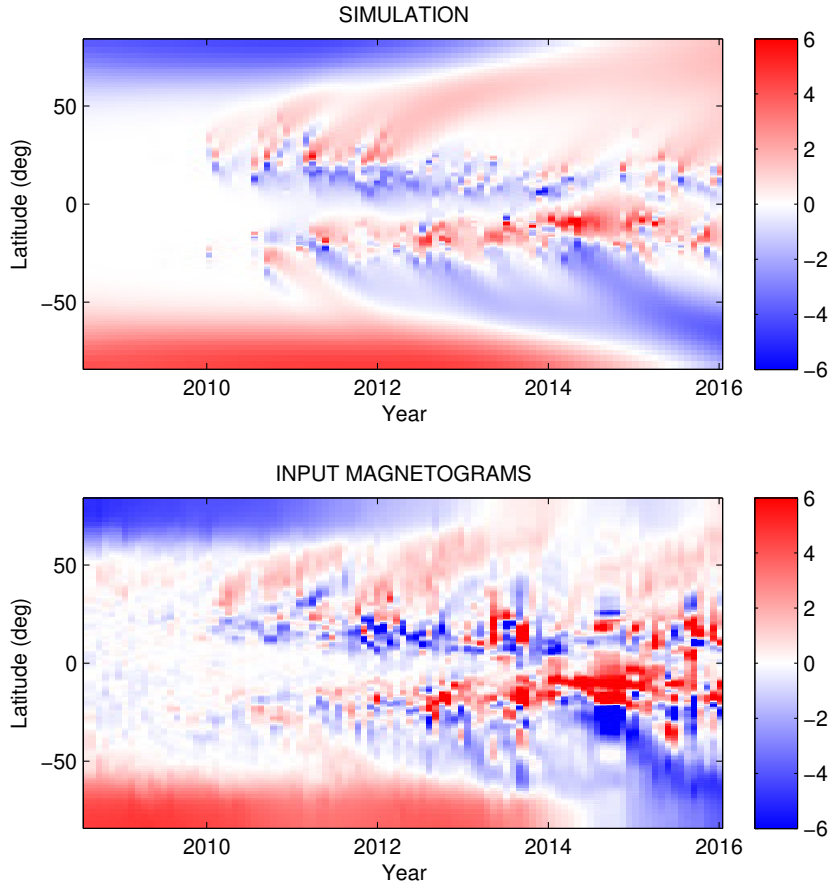

Figure 2.27: Top: Butterfly diagram for the optimal parameter 4-set for the 2D model in Table 2.5(n). Bottom: Ground truth data for Cycle 24.

2.5 Conclusions

The aim of this chapter was to use a genetic algorithm to find optimal parameters to be used for surface flux transport simulations, subsequently helping us understand the behaviour and interplay of the many physical processes on the Sun. We began by obtaining optimized parameter sets for a 1D SFT model for Cycle 23, both with and without a multiplicative tilt angle factor. From these simulations we obtained viable ranges for parameters. We found that these ranges and optimal solutions were in good agreement with results from previous studies and from observations. We also looked at the interaction of parameters, highlighting the positive correlations between the meridional velocity parameters v_0 and p, and exponential decay time τ.

We repeated the optimization process on a 2D assimilative model and found that optimum parameters were mostly within ranges of those from the 1D case, but distinct enough to suggest that the differences between models could be important. We also found an optimum value for the assimilation threshold $B_{\rm par}$, which was significantly greater than used previously for GONG magnetograms by Yeates *et al.* (2015). Qualitatively, the 2D model produced a more accurate butterfly diagram than the 1D model, particularly at the poles. We also included an exponential decay term in the 2D model which produced an optimal value of 4.5 yr, which lies outside the acceptable range found in the 1D case and is in agreement with the values obtained by other authors. Including decay induced a decrease in the velocity parameters, but given that the acceptable range extended to the upper limits of exploration, its inclusion may not be necessary in the 2D model. There is the possibility that we did not model decay realistically, which could have led to a strong polar field. That the 2D model was able to give an acceptable match to the observed butterfly diagram and axial dipole moment without a decay term is evidence that it is superior to the 1D model, which was unable to do so with the corresponding optimal parameters. It suggests that the method of flux assimilation in the 2D model is superior to the insertion of idealized BMRs, as used both in the 1D model and in most other SFT

models.

We were then able to compare the optimal meridional profiles from different regimes with observations made from feature tracking. The profiles from regimes (a), (c), and (d) were each almost completely within the range of observed flows, but the 1D optimal profile was faster than the average observed flow, while the 2D profile with decay included was too slow. The 2D profile without an extra decay term, however, best matched the average observed profile. Fixing the observed profile in both models resulted in varied success; the 2D model was able to accommodate the observations comfortably, whilst the 1D model saw a reduction in most parameters and a butterfly diagram containing an excess of flux in the transport regions.

Finally, the optimization process was repeated using the 2D model for Cycles 21, 22, and 24, producing plausible results for Cycles 21 and 22; Cycle 24 may need more time to progress to capture the long-term effects of the cycle in the optimal parameters, particularly in narrowing some of the range of viable solutions, although an optimization run performed over the same number of years for Cycle 23 showed that the optimal parameters themselves were barely affected; it was just the ranges of acceptable values which widened due to fewer constraints. In order to predict the axial dipole moment at the Cycle 24/25 minimum and hence the amplitude and length of Cycle 25, randomly generated magnetic regions with properties based on empirical relations must be used to simulate the remainder of the cycle (e.g. Upton & Hathaway, 2014*b*; Cameron *et al.*, 2016).

Analysis of multiple cycles highlighted significant differences in meridional circulation speed, supporting the evidence for slower meridional flows during stronger cycles, and initial profile strength, supporting the proposed relationship between cycle strength and polar field strength at the preceding cycle minimum. Our multiple cycle analysis also highlighted cycle-dependence of the decay term τ. At present, the best form and magnitude of such a decay term remain to be determined by the community. However, our results (and the others mentioned) do suggest that it can help to improve the match with observations, at least for Cycle 23. It is intriguing that it

seems to be less important for the preceding cycles. This could either be because the decay is compensating for some other deficiency of the model that has changed in Cycle 23, such as the inability to reproduce the unusually weak polar field at the end of the cycle, or the radial diffusion of flux did really change from one cycle to the next, presumably due to some difference in the flows and magnetic field in the convection zone. This is an interesting subject for future study, but is beyond the scope of this chapter where we consider only the surface. Radial diffusion is investigated further in Chapter 5.

All optimization runs were performed with respect to a prescribed variance which was proportional to both latitude and the observed magnetic field strength. We did this to give statistical meaning to the χ^{-2} quantity, as well as to factor in observational uncertainty. It should be noted that comparing fitness values is always with respect to the chosen error structure in this chapter.

While the flexibility in the problem is beneficial in the respect that it allows more freedom, it can also have drawbacks. For example, the choice of fitness function is crucial to deciding which regime or parameter choice is 'best' for each model, but depends entirely on what the user regards as important. Lemerle *et al.* (2015) used a combination of χ^2 statistics which measured the differences between real and simulated time-latitude maps, axial dipoles and 'transport regions' (latitudes $\pm 34°$ to $\pm 54°$). These statistics were balanced equally in the final fitness function. Weighting could have been applied in favour of particular features, though it is not obvious how best to put this into practice. Alternatively, weighting could be applied to different sections of the map, i.e., active, transport and polar regions, to force the algorithm to return parameters which produce those specific regions more accurately. We chose a comparison between the real and simulated time-latitude maps, with an associated error structure, as we considered the general reproduction of the whole map to be foremost in importance.

Chapter 3

The Bayesian approach

In Chapter 2, a parameter optimization process was performed through the use of a genetic algorithm. This method was successful when applied to a simple 1D or 2D surface flux transport model, and we would like to apply it to a more complex 3D dynamo model. However, we were running 23 000 realizations of the model to 'ensure' convergence. Whilst this was manageable for a model with a runtime of 1–2 minutes, a larger simulation that lasts a couple of hours using multiple processors would require a month of computer time for the same number of iterations, with perfect parallelization. Furthermore, as the genetic algorithm improves its current best values over time, there is no way of knowing how close one is to the optimum value, hence there is no definitive stopping criterion.

In light of this, we opt to test a different technique, known as Bayesian emulation. This method stems from the ideas of Bayesian inference, where prior beliefs are updated based on new information, and has been applied to a wide range of models from different disciplines, including oil reservoirs, galaxy formation, climate, disease, and plant biology (Craig *et al.*, 1997; Bower *et al.*, 2010; Vernon *et al.*, 2010; Williamson *et al.*, 2013; Andrianakis *et al.*, 2015; Jackson *et al.*, 2018; Vernon *et al.*, 2018). An emulator essentially mimics the behaviour of a model but is typically much faster to run. A global parameter calibration can be performed quickly and uncertainty is automatically incorporated and can be quantified simultaneously. In

addition, the emulator is able to manage difficult regions of parameter space, such as spikes and sharp gradients, or regions where the model is numerically unstable. A thorough overview of Bayesian emulation can be found in Vernon *et al.* (2010).

In this chapter we apply these techniques to the optimization problem involving the surface flux transport model in Chapter 2. In Section 3.1 we describe the set-up of the emulation framework, and our results are presented in Section 3.2. We conclude in Section 3.3.

3.1 History matching and emulation

We begin by describing the method used to perform the 'history matching' process. According to Vernon *et al.* (2014), history matching is 'modifying input parameters until the output resembles historical data'. This is in essence the same global 'parameter optimization' goal as we were aiming for in Chapter 2, but here we shall apply a different method. An emulator is described by Vernon *et al.* (2014) as a 'representation of uncertainty about the value of the model at each input choice. It suggests an approximation to the function along with an assessment of the error magnitude of the approximation'. After such approximations have been calculated, regions of parameter space can be ruled out based on a quantity called 'implausibility'.

Various levels of sophistication exist for constructing emulators. For example, one option is to fit an emulator using probability distributions for the model outputs, which would be the fully Bayesian approach (Kennedy & O'Hagan, 2001). However, here we shall use the Bayes Linear approach which only depends on expectations, variances and covariances of outputs (Goldstein, 1999; Goldstein & Wooff, 2007). Other simplifications are outlined below, but for a more complete study see Vernon *et al.* (2010).

Following the notation of e.g. Vernon *et al.* (2014), we denote model inputs (parameters) by x and outputs by $f(x)$, which correspond to properties y of the physical

system. Observations are denoted by z, and observational errors by e, so $z = y + e$, with $E(e) = 0$ and $\mathrm{Var}(e) = \sigma_e^2$, and e is assumed to be independent of y. We also consider model discrepancy ε which can be used to represent model imperfections, so $y = f(x) + \varepsilon$, with $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2$, and we specify that ε is independent of x, $f(x)$ and e.

Typically, a Bayesian emulator is constructed as:

$$f(x) = \sum_i \beta_i g_i(x) + u(x) + w(x), \tag{3.1.1}$$

where g_i are functions of the inputs x (usually low-order polynomials) which describe the global behaviour of the model, and β_i are coefficients of these functions to be determined. The second term $u(x)$ is a Gaussian process which describes the local behaviour of the model, and $w(x)$ is an uncorrelated nugget representing white noise. These have zero expectation and the following constructed covariance matrices:

$$\mathrm{Cov}(u(x), u(x')) = \sigma_u^2 \exp\left(-\frac{\|x - x'\|^2}{\Theta^2}\right), \tag{3.1.2}$$

$$\mathrm{Cov}(w(x), w(x')) = \begin{cases} \sigma_w^2 & \text{if } x = x' \\ 0 & \text{otherwise} \end{cases}, \tag{3.1.3}$$

where Θ is a specified correlation length.

For simplicity, we set $\beta_i g_i$ to be a constant term β_0. Then we have:

$$E(f(x)) = \beta_0, \tag{3.1.4}$$

and, ignoring the nugget term for simplicity:

$$\mathrm{Var}(f(x)) = \sigma_u^2. \tag{3.1.5}$$

Model outputs at inputs $x^{(j)}$ are placed into a column vector:

$$D = \left(f(x^{(1)}), f(x^{(2)}), f(x^{(3)}), \dots\right)^T, \tag{3.1.6}$$

and we write:

$$E(D) = (\beta_0, \beta_0, \beta_0, \dots)^T. \tag{3.1.7}$$

The covariance between $f(x)$ at a new input x and a known output $f(x^{(j)})$ is given by the row vector:

$$\mathrm{Cov}(f(x), D)_j = \mathrm{Cov}\left(f(x), f(x^{(j)})\right) = \sigma_u^2 \exp\left(-\frac{\|x - x^{(j)}\|^2}{\Theta^2}\right), \tag{3.1.8}$$

and the variance between two calculated outputs is given by the matrix:

$$\mathrm{Var}(D)_{jk} = \mathrm{Cov}\left(f(x^{(j)}), f(x^{(k)})\right) = \sigma_u^2 \exp\left(-\frac{\|x^{(j)} - x^{(k)}\|^2}{\Theta^2}\right). \tag{3.1.9}$$

Once these quantities have been calculated, we can then obtain the Bayes linear updated expectation and variance for the model output at any new input x, where we update our beliefs based on the actual model runs (Goldstein, 1999):

$$E_D(f(x)) = E(f(x)) + \mathrm{Cov}(f(x), D)\,\mathrm{Var}(D)^{-1}(D - E(D)), \tag{3.1.10}$$

$$\mathrm{Var}_D(f(x)) = \mathrm{Var}(f(x)) - \mathrm{Cov}(f(x), D)\,\mathrm{Var}(D)^{-1}\,\mathrm{Cov}(D, f(x)). \tag{3.1.11}$$

Equations 3.1.4–3.1.11 are quick to evaluate compared to performing a model simulation at each input. We are given an estimate of the output at each point in parameter space, along with an approximation of uncertainty given lack of information in the vicinity of that particular point. Using these quantities, we define the implausibility to be:

$$I^2(x) = \frac{(E_D(f(x)) - z)^2}{\mathrm{Var}_D(f(x)) + \mathrm{Var}(\varepsilon) + \mathrm{Var}(e)}. \tag{3.1.12}$$

The implausibility measures the difference between the emulator mean and observed data z given input parameters x (in the numerator) whilst simultaneously taking into account model and observational error, as well as the lack of emulator information in the specified region of parameter space (in the denominator). A value of $I(x) > 3$ indicates a high probability that the inputs produce an unacceptable match to observations and the corresponding regions of parameter space can be discarded (Pukelsheim, 1994).

However, our chosen output $f(x) = \chi^{-2}$ from Section 2.1.7 is already a comparison between observations and model output, subject to uncertainty, and has no 'observed value' per se. Instead we use an alternative cut-off for ruling out 'implausible' regions:

$$E_D\left(f(x)\right) + a\sqrt{\mathrm{Var}_D\left(f(x)\right)} < z_{\mathrm{cut}}, \quad a \in (2,3)\,, \tag{3.1.13}$$

where z_{cut} is chosen to be just smaller than the current highest value of $f(x^{(j)})$. This allows us to still find a region of acceptable parameter sets in the unlikely event of finding the absolute optimum from one of the simulation runs. In similar fashion to Equation 3.1.12, this equation takes into account emulator uncertainty in regions of low exploration by adding two or three standard deviations to the emulator expectation. The choice of a depends on how cautious we wish to be, or how much we trust our emulator. If the sum of the two quantities is smaller than z_{cut}, it means that the emulator is confident that a parameter input will not produce a better match to observations than the current best simulation run.

For our study we sample 100 points across parameter space using a maximin Latin hypercube (Sacks *et al.*, 1989; Currin *et al.*, 1991; Santner *et al.*, 2003). This divides the range of each parameter into 100 sub-intervals and ensures that there is exactly one point within each sub-interval. Then the hypercube with maximal minimum distance between points is chosen in order to ensure optimal coverage of parameter space. We perform full model runs at each of these points, and the emulator expectation and variance are calculated at $\sim 10^7$–10^8 other points in parameter space based on the model output from the 100 sampled parameter sets. Points below the cut-off in Equation 3.1.13 are discarded and the process is repeated in waves until the emulator variance is smaller than other uncertainties.

3.2 Applying history matching and emulation to a surface flux transport model

Rather than attempting to emulate the complex 3D dynamo model KD3 (Yeates & Muñoz-Jaramillo, 2013), we take a first step towards this by testing the process on the 2D SFT model from previous chapters (Yeates *et al.*, 2015). If we trust our optimal values in Chapter 2, we have some intuition as to where in parameter space the emulator will indicate as acceptable. However, we must ensure that this knowledge does not add bias to our search. When it comes in future to history match for KD3, we will not have the luxury of such information.

The SFT model solves Equation 2.1.1 and the free parameters are diffusion η, meridional flow parameters v_0 and p, initial dipolar field strength B_0 and assimilation threshold B_{par}. These are given the same upper and lower bounds as in Section 2.2:

(i) $100\,\text{km}^2\,\text{s}^{-1} \leq \eta \leq 1500\,\text{km}^2\,\text{s}^{-1}$

(ii) $5\,\text{m}\,\text{s}^{-1} \leq v_0 \leq 30\,\text{m}\,\text{s}^{-1}$

(iii) $0 \leq p \leq 16$

(iv) $0\,\text{G} \leq B_0 \leq 50\,\text{G}$

(v) $10\,\text{G} \leq B_{\text{par}} \leq 50\,\text{G}$

We choose our output to be $f(x) = \chi^{-2}$ as defined in Equation 2.1.25, and we expect a similar optimal value of ~ 0.65.

As discussed above, we evaluate the model at 100 maximally-spaced inputs. The respective outputs form the vector D and, maintaining the notation used in Section 3.1, we choose the prior expectation β_0 to be the mean of the entries of D. The correlation lengths Θ_i determine how strongly neighbouring points in parameter space are correlated in dimension i: a large value of Θ_i corresponds to strong correlation and vice versa. Rather than keep these fixed and equal for each of the

five parameters throughout the emulation, we estimate the values of Θ_i, as well as the standard deviation of the Gaussian process σ_u and a small random noise nugget term w, by maximizing a log-likelihood function for the normal distribution:

$$l\left(\Theta, \sigma_u, w|x\right) = -0.5\left[\log\left(\det\left(\mathrm{Var}\left(D\right)\right)\right) + \left(D - E\left(D\right)\right)^T \mathrm{Var}\left(D\right)^{-1}\left(D - E\left(D\right)\right)\right], \tag{3.2.1}$$

where $\mathrm{Var}\left(D\right)$ is the same as in Equation 3.1.9 but has been scaled by a factor of $(1 - w)$. This is in general a simple optimization problem and can be performed using the in-built interior-point algorithm in MATLAB (Byrd *et al.*, 1999). The correlation lengths, standard deviation and nugget are re-estimated at the start of every wave once new data have been obtained. However, it should be noted that these are only estimations and we must be cautious to take them at face value.

3.2.1 Waves 1 and 2

One input from the first wave of runs produces an output of $\chi^{-2} = 1 \times 10^{-6}$. This is many orders of magnitude lower than other output values and so we treat it as an outlier and discard it, leaving us with 99 runs from Wave 1. The offending input has a diffusion of $\eta = 106.9\,\mathrm{km^2\,s^{-1}}$ which causes the simulation to become numerically unstable and blow up. We find that this occurs for small values of η (with some non-linear dependence on other parameters), hence to avoid bad emulator behaviour in this vicinity we create a barrier at $\eta = 150\,\mathrm{km^2\,s^{-1}}$ and define this to be our new lower limit for the diffusivity. As discussed in Chapter 2, a more accurate timestepping method may prevent this from occurring to a larger extent.

To construct the first wave emulator we split each parameter range equally into 30 sub-intervals and calculate the emulator expectation and variance at the 30^5 points. Whilst this is not the best method for sampling inputs, it helps with visualizing the parameter space in the early stages, and in Waves 3 and 4 we shall emulate at randomly selected points.

In Figure 3.1 we plot 2D projections of the emulator expectation from 5D parameter

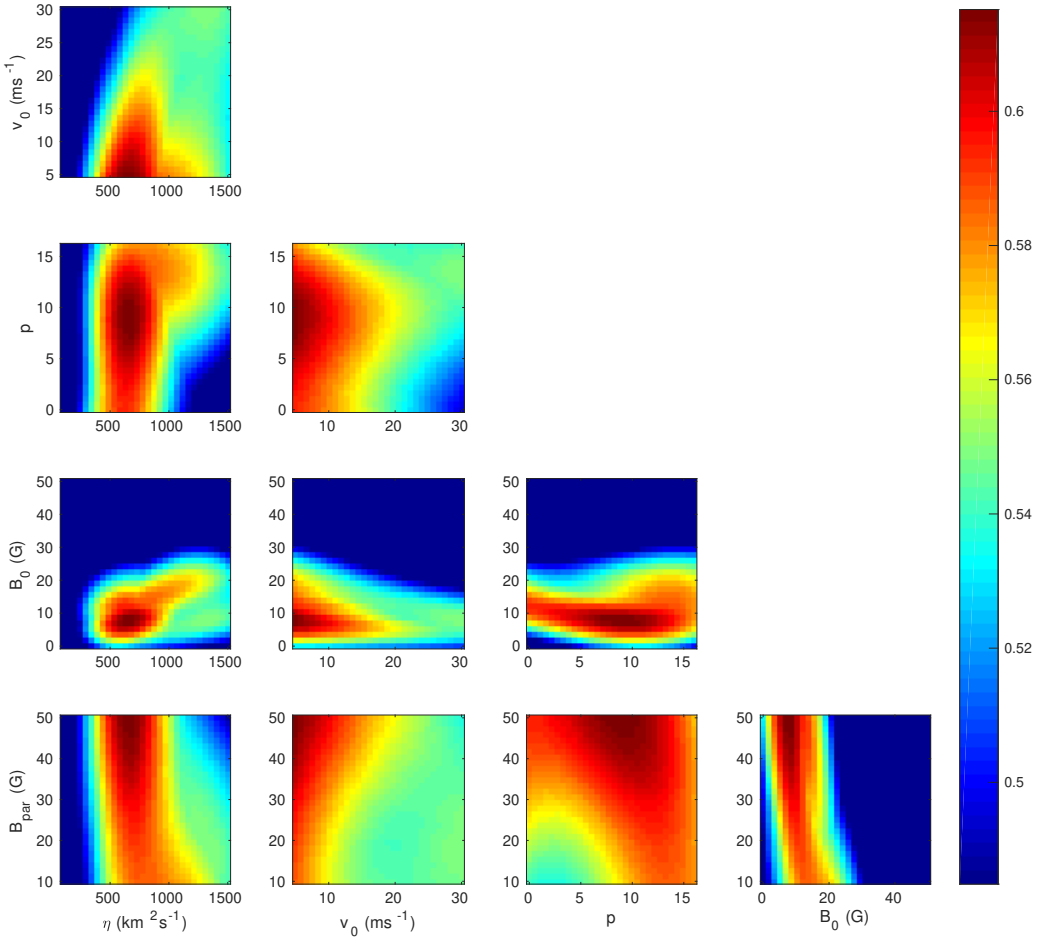

Figure 3.1: Pairs plot showing maximized expectation for each 2D projection of 5D parameter space from Wave 1.

space for each pair of parameter relationships in Wave 1. We do this by maximising the expectation over the remaining three dimensions behind each pixel. We observe some clear peaks and valleys, and are already able to infer potential correlations between parameters.

However, the emulator expectation is based purely on outputs from the 99 model runs, and there are vast regions of parameter space far away from this small sample where more information is required. Therefore we also plot our cut-off metric from Equation 3.1.13, with $a = 3$: $E_D\left(f(x)\right) + 3\sqrt{\mathrm{Var}_D\left(f(x)\right)}$, in the 'pairs plot' in Figure 3.2. This takes into account lack of information and represents the 'best-case' or 'optimistic' scenario, in that we assume there are still lots of inputs that could

produce good matches to observations. Hence in comparison with Figure 3.1, there are fewer sections which we would class as implausible, particularly for the fifth input B_{par}. This highlights the need for more model runs to gain more insight into parameter space. One striking feature is that all parameter sets with $B_0 > 30\,\text{G}$ produce a poor χ^{-2} value. This is not particularly surprising given that an initial dipole field strength that is more than twice as strong as observed would propagate throughout the cycle and the polar field would never be accurate. Despite the apparently wide ranges of acceptable input choices, we are able to discard regions below z_{cut}, which we calculate based on the current highest model output, χ^{-2}_{max}:

$$z_{\text{cut}} = \chi^{-2}_{\text{max}} \left(1 + 3\sqrt{\frac{2}{n-k}}\right)^{-1}, \tag{3.2.2}$$

where $\sqrt{\frac{2}{n-k}}$ is the standard deviation of the normalized χ^2 distribution and $(n-k)$ is the number of degrees of freedom. For Wave 1 we obtain $z_{\text{cut}} = 0.55$. This value is coloured in green in Figure 3.2. For Wave 2 we then discard all regions in 5D space below this value. This amounts to removing 87% of the original parameter space which we cannot see from the simplified visualization in Figure 3.2.

For the new sample of model runs, we generate a large number of sets of 100 parameter inputs from the remaining 13% of space, and choose the set with the maximal minimum distance to ensure good coverage of the five dimensions. After running the SFT model at the new inputs, we construct a new emulator for the second wave using only the Wave 2 points, plus any Wave 1 points which exceed the updated value of z_{cut} as defined by Equation 3.2.2. Incidentally, there are no such points from Wave 1, because the threshold is updated to $z_{\text{cut}} = 0.61$ due to the Wave 2 outputs. The second wave emulator will be more accurate in the non-implausible regions, but should not be used to approximate the discarded space.

Figure 3.3 shows the pairs plot for the best-case scenario from the Wave 2 emulator. We elect to be less cautious and speed up the process by setting $a = 2$ in Equation 3.1.13. We now observe more distinct patches of non-implausible regions, particularly

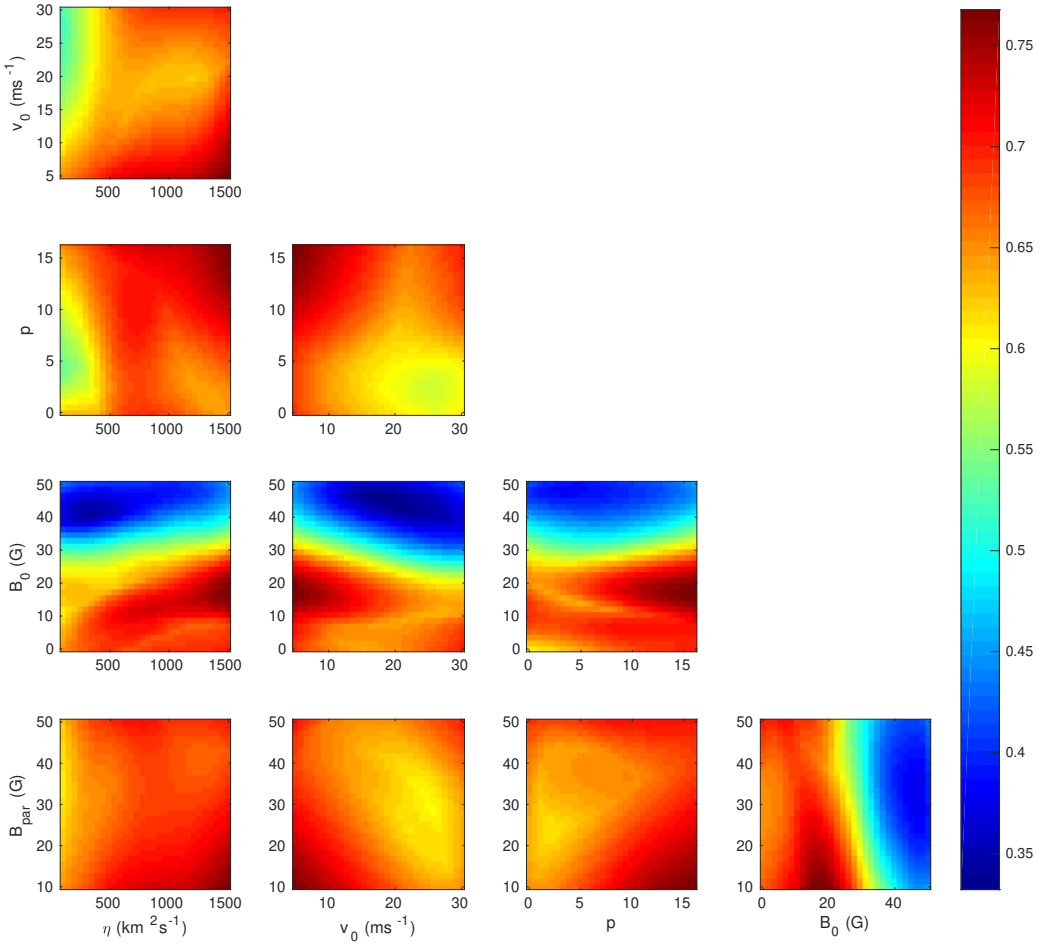

Figure 3.2: Pairs plot showing maximized best-case for each 2D projection of 5D parameter space from Wave 1.

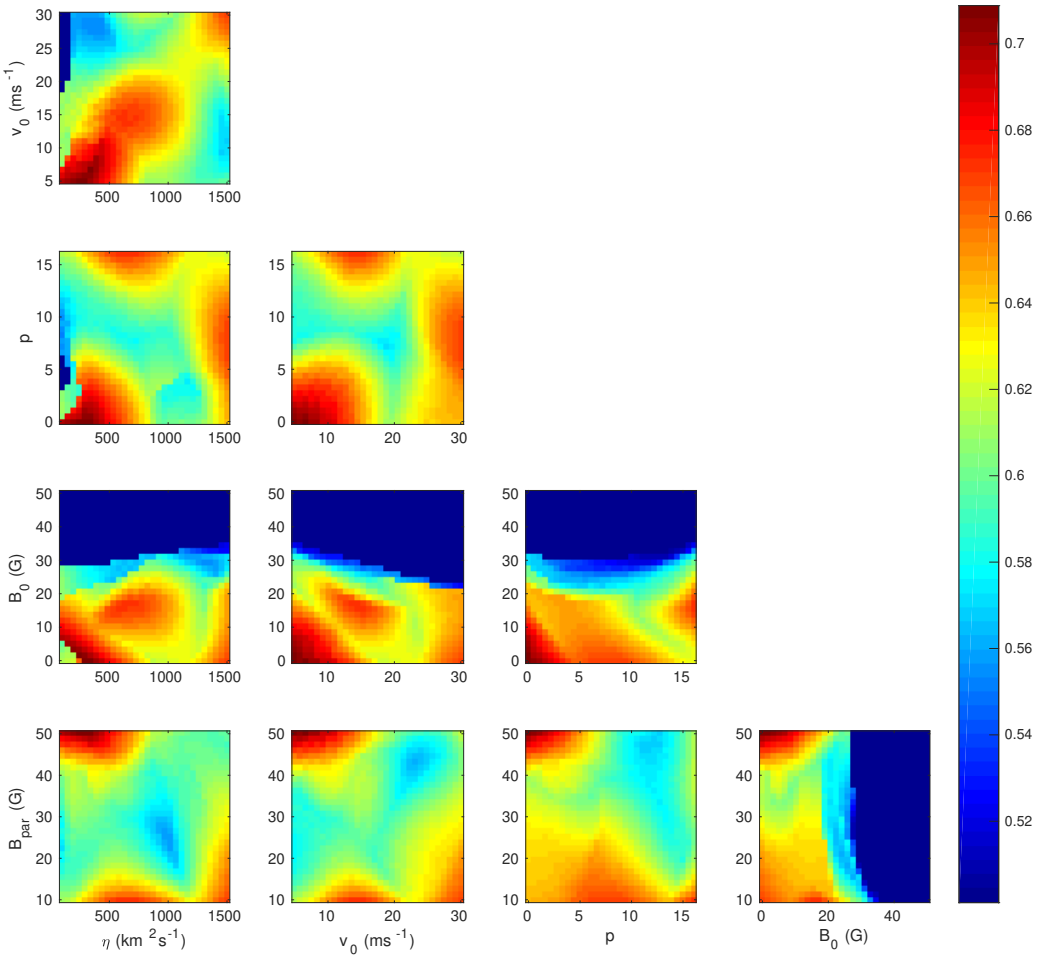

Figure 3.3: Pairs plot showing maximized best-case for each 2D projection of 5D parameter space from Wave 2.

at boundaries and corners of the 5D box, where a model run is unable to cover as large an area with its information as a model run in the centre of the domain. In time this problem will be resolved as the correlation lengths are re-estimated at each wave and should eventually ensure that boundary points are accurately emulated. We also see hints of a positive correlation between η and v_0 to add to the relationships found in Chapter 2, and that we can get erratic behaviour when small values of η are prescribed (first and second panels of the first column).

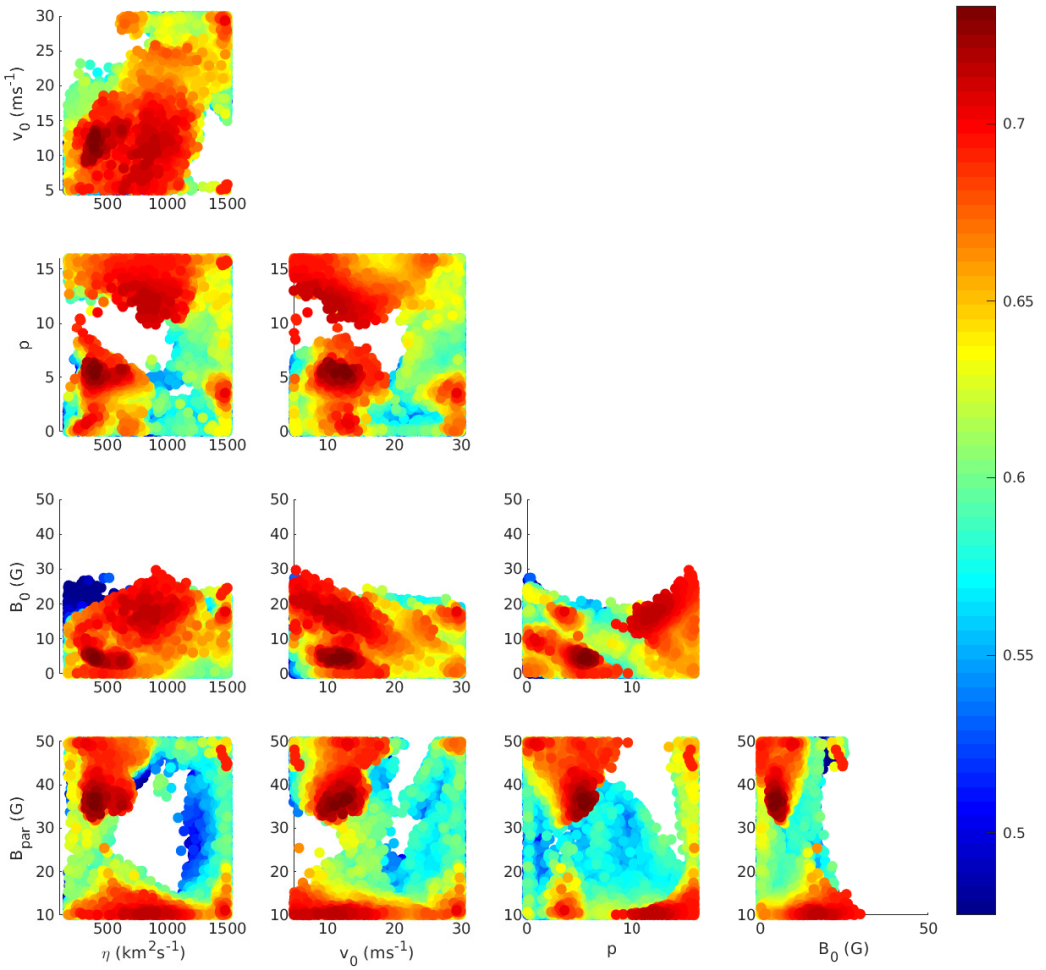

Figure 3.4: Pairs plot showing maximized best-case for each 2D projection of 5D parameter space from Wave 3.

3.2.2 Waves 3 and 4

By the time we reach Wave 3, the equally-spaced grid of 30^5 points is no longer sufficient to emulate at, especially given 99.57% of the original space has been ruled out. Instead we generate 10^7 points uniformly at random and assess their performance at each emulator level. If at any stage a point does not exceed the cut-off for the current wave, it is discarded. Conversely, if a point performs better than the z_{cut} value in Waves 1 and 2, it is passed through the third emulator based on Wave 3 runs and plotted in Figure 3.4.

We now begin to observe clear peaks in parameter space. The assimilation threshold

appears to have a local maximum at 30–40 G, and also near the lower bound of 10 G, although this could be due to boundary effects. Other parameters also appear to be bimodal, with one of the peaks in each case agreeing well with results in Chapter 2.

On reaching Wave 4, we are at the stage where we can draw 100 sparsely-distributed runs primarily from these two modes. Repeating the multi-step emulator process described above, we plot the remaining non-implausible parameter space which contains 0.11% of the original space (Figure 3.5). The history matching process has selected a clear optimal region with $\eta \in [200, 650]\,\mathrm{km^2\,s^{-1}}$ and the very best values of η falling between 300–450 $\mathrm{km^2\,s^{-1}}$. The maximum meridional flow speed v_0 produces a good match when values between $7\,\mathrm{m\,s^{-1}}$ and $14.5\,\mathrm{m\,s^{-1}}$ are prescribed, in excellent agreement with Chapter 2 (see Table 2.5(c) for comparison). We again find that choosing $p < 5$ is required to match observations, although in this case values of $p = 0$ are also deemed acceptable. In this scenario, Equation 2.1.24 reverts to a simple cosine curve, indicating a lack of dependence on the exact shape of the meridional flow profile, provided the peak velocity is not too close to the equator. This is the largest deviation from the results of Chapter 2. An acceptable choice for B_0 lies in the range 3–15 G, although the current best value for B_0 is slightly higher than the 8.3 G found by the genetic algorithm. The assimilation threshold B_{par} can still take a wide range of values, but seems to be cut off below 30 G. The upshot is that the results from the emulation exercise are qualitatively, and in most cases quantitatively, in good agreement with the results obtained using `PIKAIA`.

However, there is also a considerable population with a best-case value of approximately 0.63, represented by the large blue-green mass corresponding to, for example, high values of p. Furthermore, there are individual points (most clearly visible in the B_0 vs η panel) that supposedly produce a good fit to observations despite being far away from the main population. We cannot yet rule these out, because a high best-case value indicates either a location in parameter space with little information, or a point where a model run with a high χ^{-2} has been found. At this stage we no longer need to run another 100 iterations of the model; we can simply run the

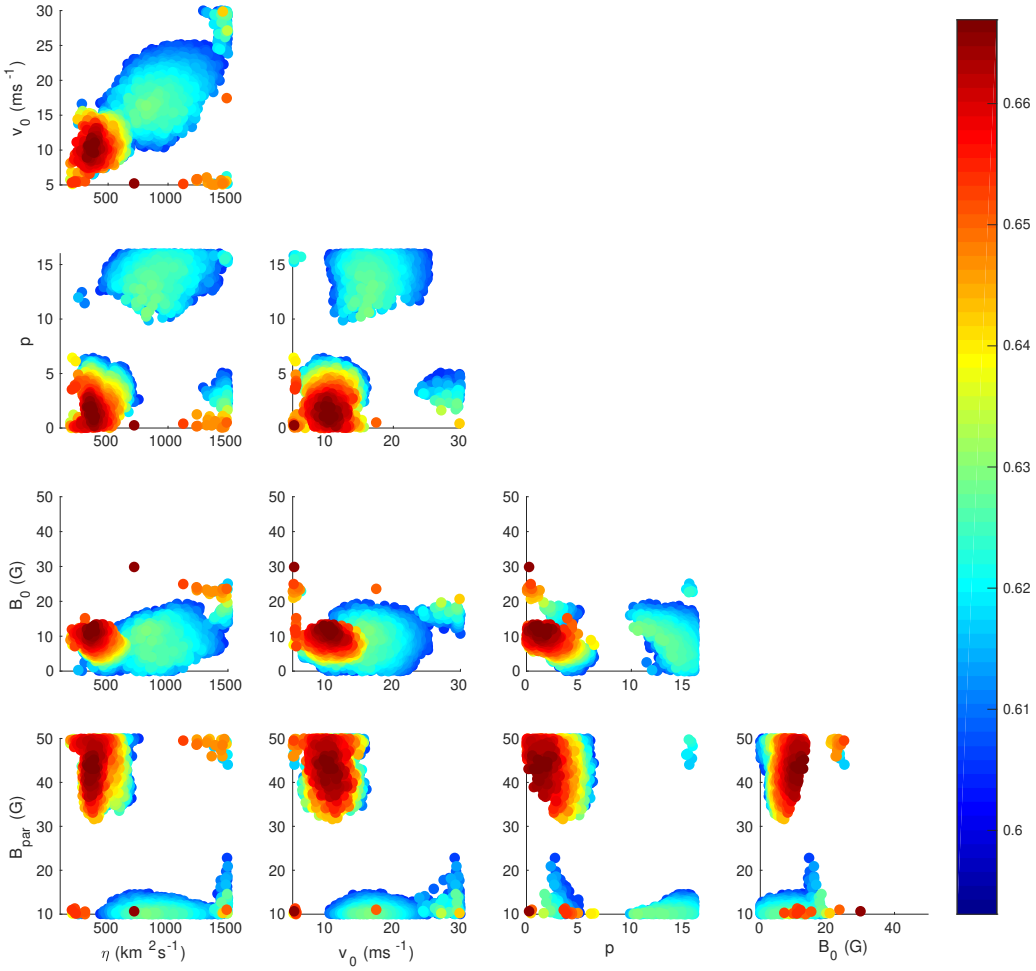

Figure 3.5: Pairs plot showing maximized best-case for each 2D projection of 5D parameter space from Wave 4.

simulation at the specific locations of interest to determine their χ^{-2} values directly. We find that these are points with large emulator variances as a result of being at the boundaries of the 5D space, and confirm that the χ^{-2} values are less than 0.5. The highest χ^{-2} value in the green secondary peak is 0.58 – this is modest but not as high as the model runs performed at the primary peak ($\chi^{-2}_{\text{max}} = 0.648$). An alternative and more sophisticated method would be to construct a separate emulator for the secondary hill, but due to time constraints and confidence in our current emulator, we decide against this. After Wave 4, 99.98% of the original parameter space is discarded.

For direct comparison, in Figure 3.6 we show another pairs plot but for the population generated by the genetic algorithm `PIKAIA` in Section 2.2. Note that the colour scale is different for improved visualization. We see that the optimal regions of parameter space obtained by the genetic algorithm and the emulator agree very well, as well as the fact that large values of B_0 produce poor fits to the ground truth data. Unfortunately, we see that `PIKAIA` does not provide information for the whole of the 5D parameter space, and while it is a global optimizer, it rapidly focuses in on the optimum, leaving large portions of space undiscovered. We also observe horiztonal and vertical lines in a grid-like structure, showing how the genetic algorithm jumps about parameter space slowly by typically only varying one parameter value at a time via mutation or crossover.

3.2.3 Next steps

In spite of the excellent progress and promising results, we would like to continue to precisely map out the shape of this probable global maximum peak. To do this we perform 100 well-spaced new model runs at points in the non-implausible space with highest emulator variance. Similarly, we choose 10–20 points with high emulator expectation, and append these runs to the Wave 4 outputs. This is ongoing work, but eventually we hope to have a much clearer idea about the behaviour of the model

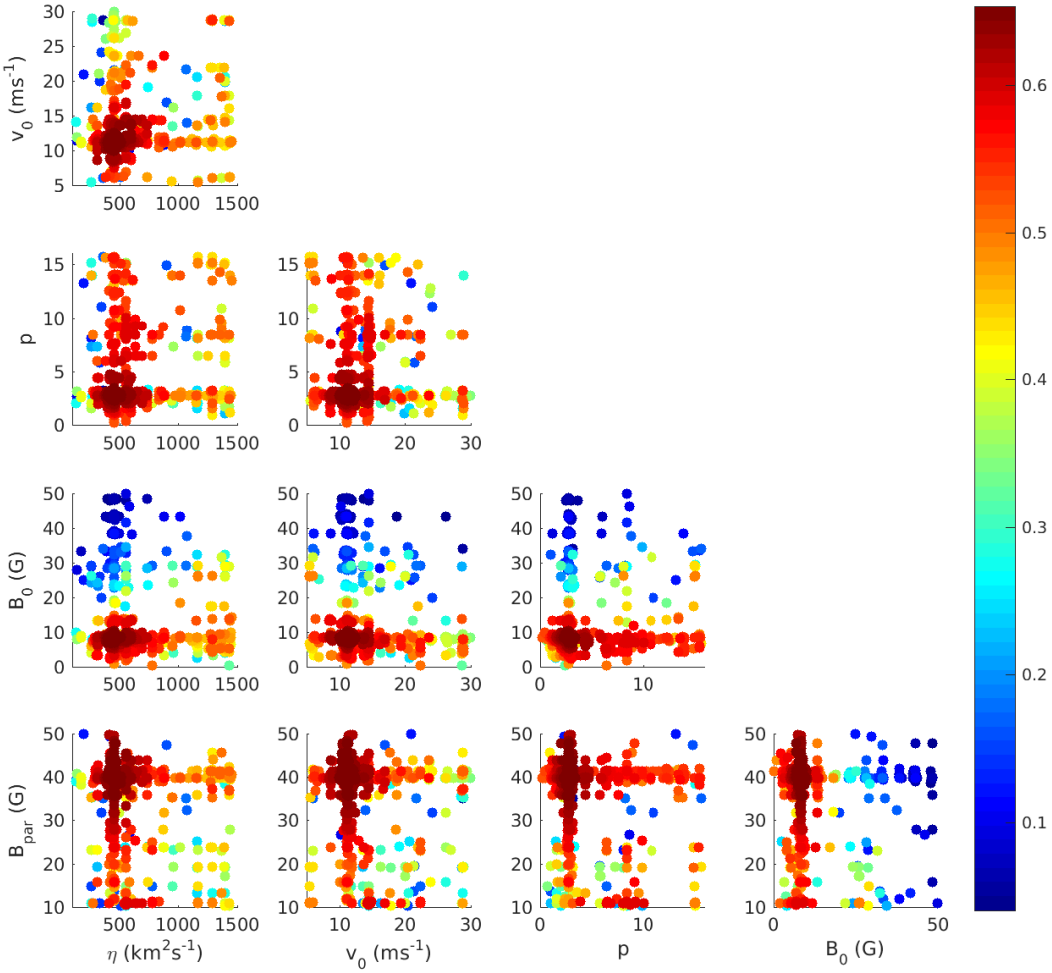

Figure 3.6: Pairs plot showing χ^{-2} for each 2D projection of 5D parameter space generated by the genetic algorithm in Section 2.2.

and its parameters in this region of parameter space, and finally obtain 'acceptable ranges' of parameters to be used in the SFT model to produce sun-like output. These ranges can also be compared closely to those found in Chapter 2.

3.3 Conclusions

In this chapter, we repeated the parameter optimization task from Chapter 2 by applying Bayesian emulation to the 2D SFT model of Yeates *et al.* (2015). Based on just 100 model runs, the emulator provided an approximation of the model's ability to match observations at each point in 5D parameter space, along with an estimate of the error of the approximation. This is far more efficient than using a genetic algorithm, because it is much faster to evaluate the emulator at a point in parameter space than to run the model. Parameter inputs with a high chance of producing poor fits to data were discarded.

Using ideas stemming from Bayesian inference, we were then able to update and improve our prior understanding of the parameter space by performing 100 more model runs in the remaining space. This process was repeated in four waves, until we had eventually honed in on an optimal region in 5D space. The parameter ranges agreed qualitatively very well with the acceptable ranges found in Section 2.2. A final step will be to perform one more wave of runs in the optimal region to gain a better picture of the global maximum peak.

Model runs were still required to develop the emulator, but we only needed to run 400 to reduce the parameter space to that shown in Figure 3.5. It might even be the case that we can obtain the same optimal region by only performing, say, 200 simulations. Contrast this with the genetic algorithm, where 23 000 realisations of the model were performed to ensure convergence. This also raises a further benefit of the Bayesian approach: the whole parameter space and the current status of the emulator can be visualised regularly. After each wave, we can track how much space is being discarded, assess the distribution of emulator expectation and variance, and

update our choice of cut-off and correlation lengths. After Wave 1, 87% of space was immediately cut, and after Wave 4, 99.98% of the original space had been ruled out, highlighting the rapid convergence of the parameter search. Furthermore, using emulation we were able to identify secondary peaks and local maxima and choose whether to explore them further. This was not possible using the genetic algorithm.

As mentioned in Chapter 2, in the future we would like to optimize parameters for the more complex 3D dynamo model, KD3. Based on the work carried out in this chapter, Bayesian emulation looks to be a more efficient and informative method than using a genetic algorithm, with opportunities to add even more levels of efficiency and sophistication. Nevertheless, it will inevitably still be a time-consuming operation and we must leave this for future work.

Chapter 4

How many active regions are necessary to predict the solar dipole moment?

Having optimized parameters for SFT models, we now turn our focus to solar cycle prediction and the role of active regions in the Babcock-Leighton mechanism. We aim to discover how much individual regions can contribute to the polar field at cycle minimum, with particular interest in the unusually weak polar field (and equivalently weak axial dipole moment) at the end of Cycle 23 (Muñoz-Jaramillo *et al.*, 2012), which in turn is believed to be responsible for the low amplitude of Cycle 24. For the published version of this chapter, see Whitbread *et al.* (2018).

Jiang *et al.* (2015) used the BMR data of Li & Ulrich (2012) to investigate the effect of tilt angle on the asymptotic axial dipole moment contribution D, using an empirical relation involving tilt angle, latitude and area (Jiang *et al.*, 2014):

$$D \propto A^{\frac{3}{2}} \sin\alpha \, \exp\left(-\frac{\lambda^2}{110}\right), \tag{4.0.1}$$

where A is the area, α is the tilt angle, and λ is the latitude of each active region at the time of emergence. They found that axial dipole moment contributions from observed tilt angles in Cycle 23 follow those obtained by assuming Joy's Law at

latitudes above ±10°. Nearer the equator, the regions with observed tilt angles contribute substantially less than would be expected from Joy's Law, contrary to the behaviour of Cycles 21 and 22, which follow the Joy's Law contributions more closely at all latitudes. This led to the suggestion that a single large anti-Hale or anti-Joy region emerging at a low latitude, or across the equator (Cameron *et al.*, 2013, 2014), has the ability to significantly alter the dipole moment, and this could have been the catalyst behind the weak polar field at the end of Cycle 23. Therefore the stochasticity behind the properties of emerging regions provides a problem for those attempting to predict the amplitude of future cycles, especially given that the magnetic flux in a single large active region is similar to the total polar flux (Wang & Sheeley, 1991). With this in mind, it may not be possible to make reliable predictions until the end of the cycle, unless random fluctuations of active region properties are taken into account. Indeed, Nagy *et al.* (2017) recently demonstrated in a 2×2D dynamo model that large 'rogue' regions can drastically affect the evolution of future solar cycles and introduce hemispheric asymmetries. Such large regions emerging during the early phases of a cycle can even affect the amplitude and duration of the same cycle. In this particular dynamo model, the effect of a single region can persist for multiple cycles. Nagy *et al.* (2017) found that the effect of a region in their model is dependent on its axial dipole moment at time of emergence. So bipolar regions near the equator, and/or with large tilt angle, are particularly strong contributors, although significant effects were found for regions even up to ±20° latitude.

We investigate these claims further by simulating the evolution of real active regions from Cycles 21 to 24 using the 2D SFT model from Section 2.2. The automated region identification and assimilation process allows us to identify particular observed properties which could have defined the contribution of each region to the axial dipole moment. As in Section 2.2, the emerging regions are determined from NSO line-of-sight magnetograms. In Section 4.1 we discuss the extraction of regions and their properties in more detail. In Section 4.2 we show how assimilating different numbers of regions based on both dipole moment contribution and flux can alter the

end-of-cycle axial dipole moment. In Section 4.3 we investigate in more detail how the properties of the regions determine their dipole contributions, and in Section 4.4 we assess the effect of the exponential decay term. We conclude in Section 4.5.

4.1 Determination of active region properties

We will investigate the distribution of various magnetic region properties, namely latitude, magnetic flux, and initial and final axial dipole moment. The regions and their properties are extracted from NSO Kitt Peak and SOLIS synoptic magnetograms, and the overall photospheric evolution is simulated using the 2D SFT model described in Section 2.2 and Yeates *et al.* (2015). The extracted active region data are freely available at the Solar Dynamo Dataverse[a] (Whitbread, 2019). All simulations are performed using optimal values for diffusivity, meridional flow, initial field strength, exponential decay and assimilation threshold, obtained using `PIKAIA`. In Section 4.4 we present the case without decay and show that similar conclusions hold in both regimes. The optimum values are shown in Table 4.1, with associated 'acceptable ranges' below each entry. We use the same value of $B_{\text{par}} = 39.8\,\text{G}$ as found in Section 2.2. All conclusions made in this Chapter are with respect to these optimal parameter values. Note that these differ slightly to those in Table 2.5, because we keep the parameters fixed across Cycles 21 to 23 and re-perform the optimization against the butterfly diagram provided by Petrie (2012). Unfortunately, the ground-truth data only goes up to 2016 Jan 1, so for Cycle 24 we assume that the optimal parameters from earlier cycles are appropriate for use, given that they lie within the acceptable ranges for Cycle 24 in Table 2.5(n).

Note also that B_0 is the initial field strength at the start of Cycle 21 only; each other cycle immediately follows on from the final state of the preceding cycle. As mentioned in Section 2.5, if we calibrate the parameters for each cycle individually, we find that any errors in the end-of-cycle dipole moment induce errors in the initial

[a] https://dataverse.harvard.edu/dataverse/solardynamo

η ($\mathrm{km^2\,s^{-1}}$)	v_0 ($\mathrm{m\,s^{-1}}$)	p	τ (yr)	B_0 (G)
466.8	9.2	2.33	10.1	6.7
[325.7, 747.3]	[5.6, 11.9]	[1.12, 3.95]	[3.6, 31.9]	[0.0, 15.0]

Table 4.1: Optimal parameter set for the simulation shown in Figure 4.1. Upper and lower bounds for acceptable parameter ranges are given in square brackets below each entry, although here we use the optimum values themselves for all simulations.

state of the next cycle, and these discrepancies can propagate through, in some cases forcing parameters of future cycles to be unrealistic. Keeping the parameters fixed throughout ensures that both the overall magnetic field and dipole moment are well reproduced, and validates any comparisons made between cycles. The optimal butterfly diagram for Cycles 21 to 24 is shown in the top panel of Figure 4.1, and the bottom panel shows the observed butterfly diagram from Kitt Peak, also partly shown in Figure 1.6.

The axial dipole moment of region i is given by:

$$D^{(i)}(t) = \frac{3}{2}\int_0^{\pi}\int_0^{2\pi} B^{(i)}(\theta,\phi,t)\cos\theta\sin\theta\,d\phi\,d\theta, \tag{4.1.1}$$

where $B^{(i)}(\theta,\phi,t)$ is the evolving magnetic field of the individual region i, computed after its initial insertion by solving Equation 2.1.1 with no other field present. Isolating the evolution of a single region like this is meaningful because Equations 2.1.1 and 4.1.1 are approximately linear, so that the contributions $D^{(i)}(t)$ may be added together to give the overall dipole moment $D_{\mathrm{tot}}(t)$. The linearity is only approximate because our newly inserted regions replace pre-existing flux, and in some cases, very strong regions can reappear in the magnetogram of the following Carrington rotation. Because of complex flux emergence and cancellation processes that occur between the multiple observations of the same region, it is not trivial to automatically define whether an active region is new or a repeat in the model, so we class these repeats as new regions altogether. The replacement technique ensures that the axial dipole moment contribution from a returning region is not counted

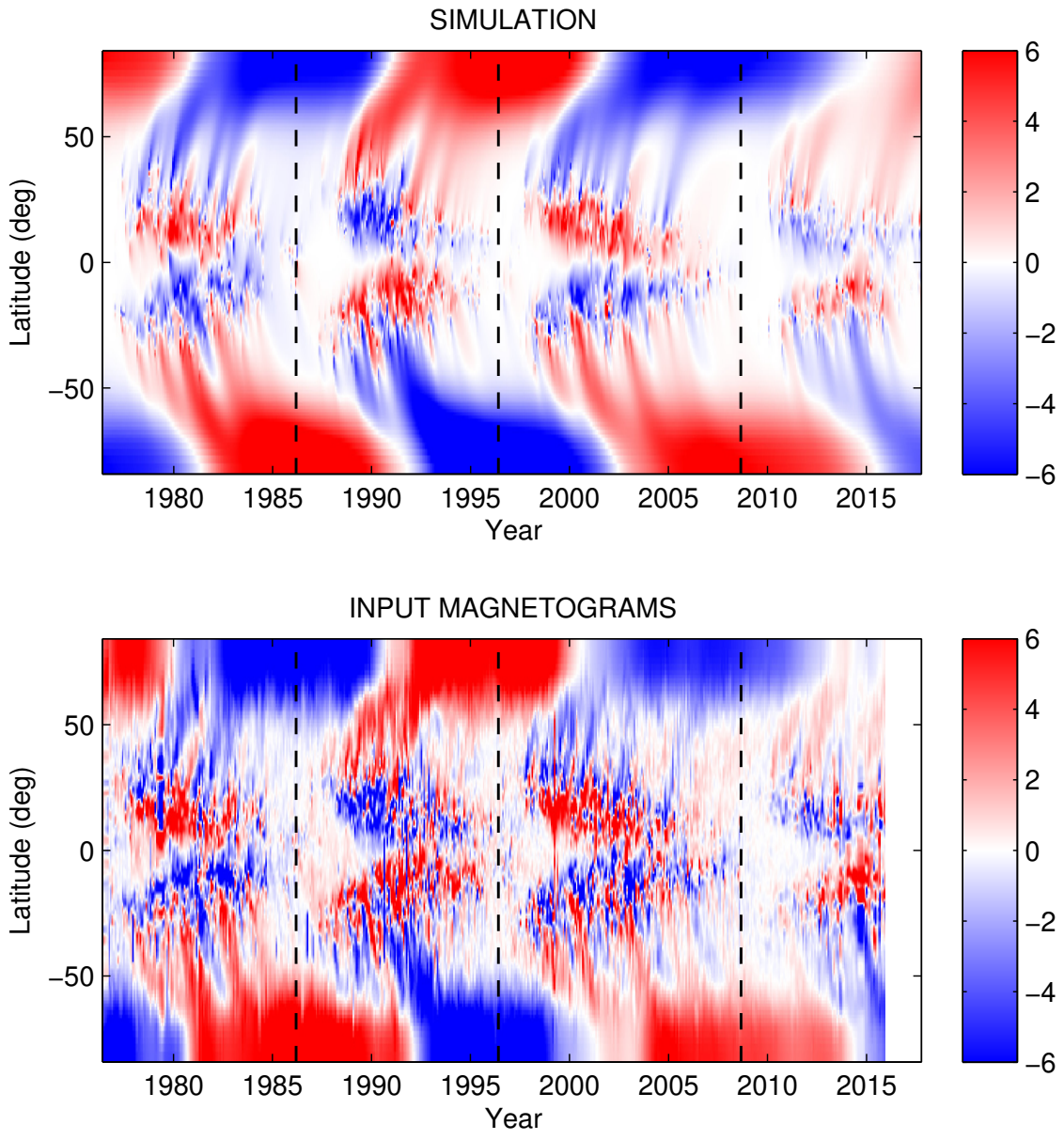

Figure 4.1: Top: Optimal butterfly diagram for Cycle 21 through to Cycle 24, simulated using the parameters from Table 4.1. Bottom: 'Ground truth' data for the same period. Vertical dashed lines indicate start/end points of cycles as used in this thesis.

twice, hence the repeated regions do not affect our conclusions. In any case, the evolution of the strongest of a set of repeated regions is a good approximation to the combined evolution including replacements, and it is therefore useful to isolate them.

To assess the contribution of each region to the overall evolution of the dipole moment, we will also use the relative axial dipole moment D_{rel}, which is defined as:

$$D_{\mathrm{rel}}^{(i)}(t) = \frac{D^{(i)}(t)}{D_{\mathrm{tot}}(t_{\mathrm{end}}) - D_{\mathrm{tot}}(t_{\mathrm{start}})}, \tag{4.1.2}$$

for region i, where $D_{\mathrm{tot}}(t)$ is the dipole moment of the full simulation with all regions included, and $D^{(i)}(t)$ is the dipole moment contribution of a single active region as calculated in Equation 4.1.1. The times t_{start} and t_{end} are the start and end of each cycle respectively, so that $D_{\mathrm{rel}}^{(i)}$ represents the contribution from region i to the overall change in dipole moment during the cycle. The start and end times are the same as in Chapter 2, i.e.: t_{start} = 1976 May 1 and t_{end} = 1986 March 10 for Cycle 21, t_{start} = 1986 March 10 and t_{end} = 1996 June 1 for Cycle 22, and t_{start} = 1996 June 1 and t_{end} = 2008 August 3 for Cycle 23. For Cycle 24 we choose t_{start} = 2008 August 3 and t_{end} = 2017 November 6. The final relative axial dipole moment $D_{\mathrm{rel}}^{(i)}(t_{\mathrm{end}})$ then reflects the proportional contribution of region i to the end-of-cycle axial dipole moment. A positive $D_{\mathrm{rel}}(t_{\mathrm{end}})$ corresponds to a strengthening of the axial dipole moment at the end of the cycle, whilst a negative $D_{\mathrm{rel}}(t_{\mathrm{end}})$ corresponds to a weakening.

Note that most SFT simulations, including Jiang *et al.* (2015), assume that all regions are BMRs with a simple bipolar structure. However in our 2D model this is not always the case. The model inserts the observed shapes of active regions, meaning that complex multipolar configurations are often assimilated. Figure 4.2 shows the configurations of the top nine largest contributors from Cycle 23, as measured by $D_{\mathrm{rel}}(t_{\mathrm{end}})$. Among these are two regions that share similar features (left and centre panels of the middle row), and are likely to have been the same region appearing in two consecutive rotations, having undergone some sort of interaction

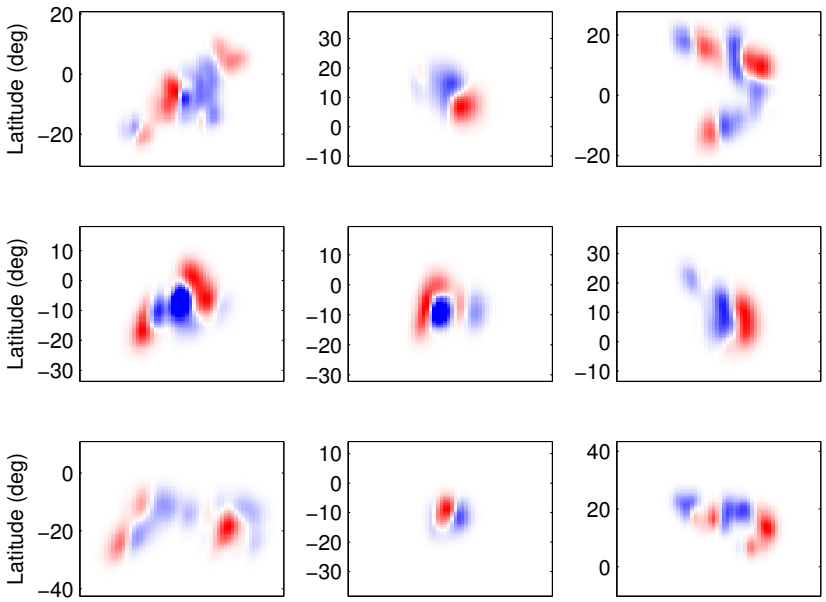

Figure 4.2: Nine most significant contributing regions from Cycle 23, as measured by $D_{\mathrm{rel}}(T)$. The panels are equal in size and centred around each region. Each image is saturated individually.

in the interim. Whilst some regions are clearly bipolar, some are less clear and are harder to separate into BMRs. Because of this, a 'tilt angle' is no longer a sensible measure, and so instead we use the initial (relative) axial dipole moment which still takes into account orientation and polarity. Similarly, we also do not consider polarity separation distance. Here the initial axial dipole moment of an active region is measured at the time of assimilation, that is, on the day it crosses the central meridian.

For the optimal threshold B_{par}, we tend to extract fewer regions per cycle than other studies, because the model can consider a cluster of active regions to be one single large region. Despite this, the insertion of realistic configurations of active regions combined with the optimization procedure means that the evolution of the observed axial dipole moment D_{tot} is well reproduced by the simulation, even though the axial dipole moment is not considered directly in the fitness function (unlike Lemerle *et al.*, 2015). We will also continue to use the term 'regions' to describe both individual

regions and clusters of regions.

4.2 How many regions are required?

Initially we consider the effect on the overall axial dipole moment of including the largest dipole moment contributions only, to assess how many regions are needed to replicate the original axial dipole moment. Regions are listed in order of absolute $D_{\mathrm{rel}}\left(t_{\mathrm{end}}\right)$ and only those above a certain threshold are assimilated. This routine is performed at five thresholds so that the top 10, 100, 250, 500 and 750 regions are included over five separate runs in each cycle, and the resulting profiles are shown in Figure 4.3(a). These are superimposed on the observed axial dipole moment (light grey). Because we only analyse a portion of Cycle 24, which was in itself a less active cycle, we instead consider the effect of including 10, 100, 200, 300 and 400 active regions.

The left-hand section of Figure 4.3(a) shows the effect of keeping the largest contributions to the axial dipole moment from the simulation of Cycle 21. Incorporating the largest 750 contributors of the 844 regions makes only a little difference (a decrease of 1.6%), but using 500 regions corresponds to a reduction of 7% of the axial dipole moment.

The left-middle section of Figure 4.3(a) shows the effect of including the largest contributions to the axial dipole moment from the simulation of Cycle 22. As few as 500 of the 846 regions can be used with a shortfall of just 1.3%, and using 750 regions makes little difference to the evolution of the axial dipole moment. If we assimilate the top ten contributors of Cycle 22, polar field reversal is almost achieved.

The right-middle section of Figure 4.3(a) shows the same profiles as the two leftmost sections but for Cycle 23. Even when the largest 750 contributors of the 951 regions are assimilated, there is a more significant discrepancy (a decrease of 4.7%) between the resulting axial dipole moment and D_{tot} than in the previous two cycles. We

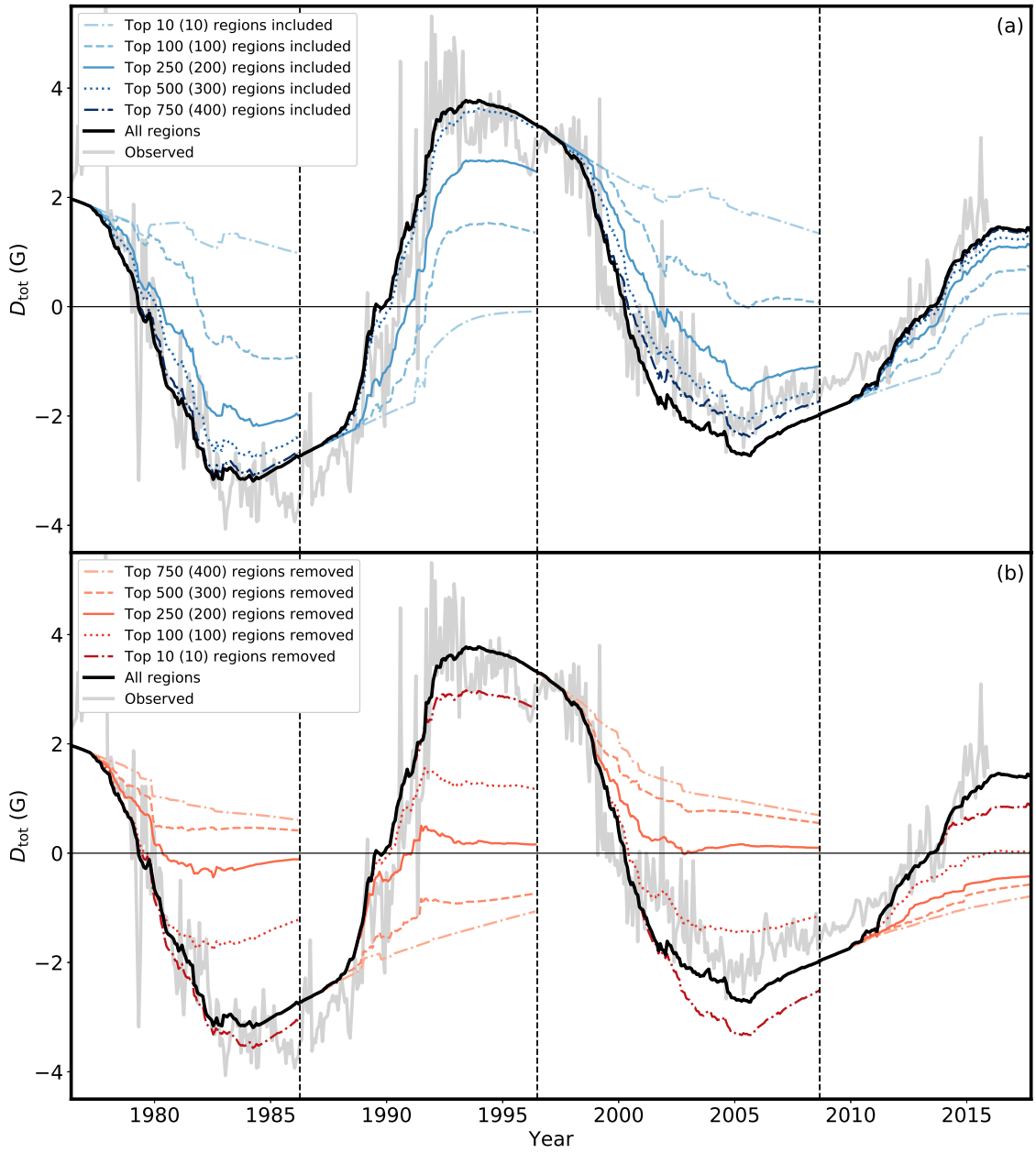

Figure 4.3: Evolution of the axial dipole moment for Cycles 21 to 24. Each profile is obtained by: (a) only using a certain number of the biggest contributors to the axial dipole moment, or (b) removing the biggest contributors to the axial dipole moment. Numbers in brackets apply to Cycle 24 only. Colour intensity is indicative of the number of regions used in each simulation, as shown in the legend. The light grey curve shows the observed axial dipole moment. Vertical dashed lines indicate start/end points of cycles as used in this thesis.

will show later that this is because most of the large contributors in Cycle 23 act to weaken the overall dipole moment (opposite to the majority pattern). The cumulative contribution of many weaker regions is therefore needed to recover its final strength. So although a small number of regions have a disproportionate effect, the cumulative contribution of the many regions with weaker dipole moment cannot be ignored, owing to their common sign.

The right-hand section of Figure 4.3(a) shows the effect of including the largest contributions to the axial dipole moment the simulation of Cycle 24. This is a weak cycle, so as few as 200 regions are required to obtain a good match to the original simulation. The quantitative behaviour of the profiles is similar to that of Cycle 22, particularly the ability to almost achieve polar field reversal with just 10 regions.

In each cycle we see that the top $\sim 10\%$ of contributors (that is, about 100 of them) determine the rapid short-term changes in the axial dipole moment. Here we see the deficit in Cycle 23; even when the top 100 contributors are included the polar field is still unable to reverse. If we remove the top 10 strongest regions from the simulation instead of keeping them (Figure 4.3(b)), we discover that the amplitude of the final axial dipole moment is overestimated in Cycles 21 and 23, and underestimated in Cycles 22 and 24. This demonstrates the impact of the strongest regions from the four cycles, and that the polar field at the end of Cycle 23 could have been stronger had the strongest few regions emerged with different properties or not emerged at all. If the top 100 strongest regions are removed from Cycle 23, the axial dipole moment is better represented than in the equivalent cases for Cycles 21 and 22, presumably because the proportion of regions with negative dipole moment contribution is greater in Cycle 23.

It is imperative to note that there are more total regions involved in our simulations of Cycle 23 than in Cycles 21 and 22, because it is a weaker and therefore longer cycle. The consequence of this is that the same number of regions in Cycle 23 represents a smaller proportion of the total number of regions compared to the other two cycles, and so naturally we might expect the axial dipole moment to be weaker when using,

Cycle	10%	20%	30%	40%	50%	60%	70%	80%	90%
21	11.3	7.31	5.36	4.04	2.85	2.06	1.48	0.97	0.53
22	14.8	9.17	6.62	4.60	3.30	2.37	1.59	1.06	0.60
23	12.9	7.98	5.21	3.78	2.79	2.08	1.51	0.98	0.49
24	8.66	5.68	4.32	3.22	2.54	1.95	1.39	0.99	0.48

Table 4.2: Flux thresholds corresponding to the top x% of regions per cycle when ordered by flux. Values given are multiples of 10^{21}.

say, 750 regions in Cycle 23. However, when we balance the proportion of regions with the previous cycles, there is still a larger difference between the profile with all regions and the profile with some regions removed. We conclude that ultimately this pattern comes down to the polarity distribution of regions with a small contribution to the axial dipole moment, and that the smallest 100 contributors of Cycle 23 must have predominantly positive $D_{\mathrm{rel}}(t_{\mathrm{end}})$.

4.2.1 What are the implications for making predictions?

Up to this point regions have been ordered by $D_{\mathrm{rel}}(t_{\mathrm{end}})$. Unfortunately, calculating this at time of emergence requires us to know the subsequent behaviour of all other regions during the rest of the cycle, which is not ideal if the aim is to predict future contributions and the required information is not yet available. Therefore we now examine the consequences of ordering and including regions based on absolute flux, which is a quantity readily measured at time of emergence. The solid lines in Figure 4.4 display the change in $D_{\mathrm{rel}}(t_{\mathrm{end}})$ as more active regions are included in the simulation, ordered by decreasing flux, for Cycles 21 (pink), 22 (yellow), 23 (dark green) and 24 (brown). Table 4.2 shows some examples of flux thresholds.

There are multiple regions with large flux that contribute positively to the axial dipole moment during Cycle 21. Because of this, 80% of $D_{\mathrm{tot}}(t_{\mathrm{end}})$ is attained when less than 40% of regions are considered (bearing in mind the threshold for the top 40% is ~ 3.5–4.5×10^{21} Mx depending on the cycle). There is then a sharp decrease

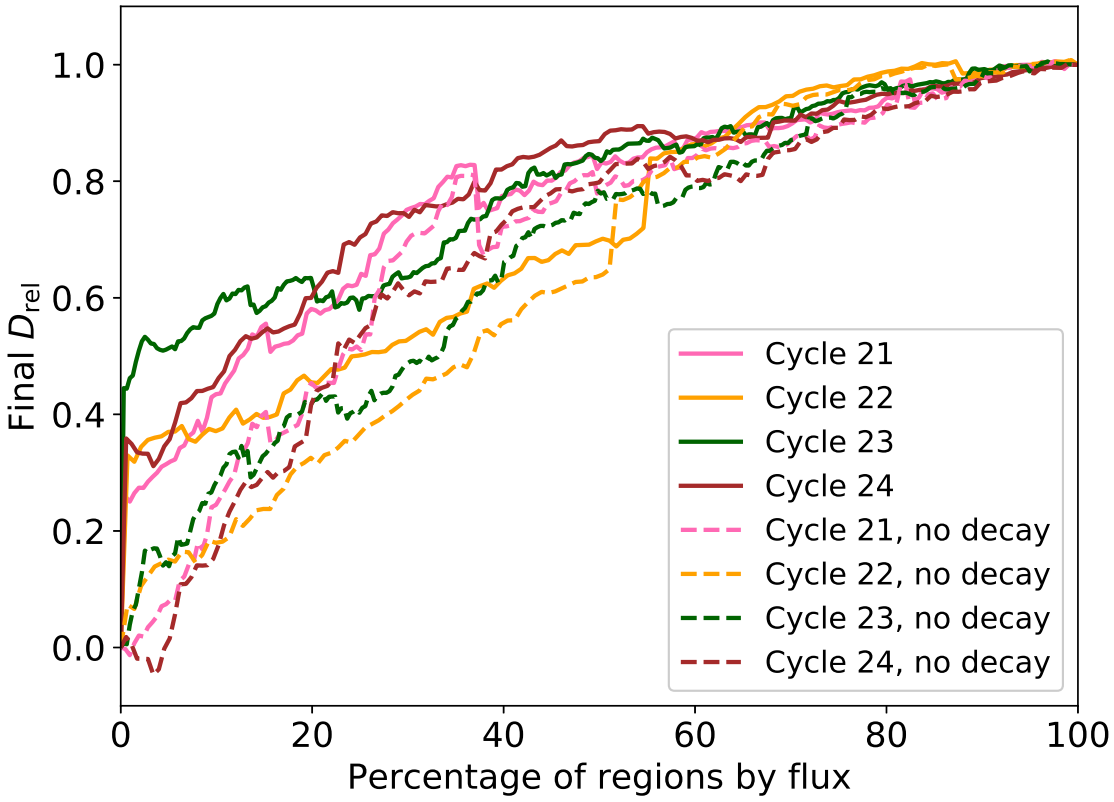

Figure 4.4: Final D_{rel} against percentage of regions included for Cycles 21 (pink), 22 (yellow), 23 (dark green) and 24 (brown). Solid lines are the cases with exponential decay, and dashed lines are the cases where the decay term has been removed. Regions are ordered by flux and the top $x\%$ of the strongest regions are incorporated. Some thresholds for inclusion are given in Table 4.2.

when the two biggest contributions of $D_{\mathrm{rel}}(t_{\mathrm{end}})$ are included, before the 80% mark is reached again, corresponding to half the number of regions being used. Note that more than 25% of $D_{\mathrm{tot}}(t_{\mathrm{end}})$ is attained by using only a small percentage of the largest regions. This is a side-effect of the measure we use. For example, when decay is not present (see Figure 4.9 in Section 4.4) and 10 regions are included, the end-of-cycle dipole moment is far away from the original end-of-cycle dipole moment (thick black line), and the contribution is small (dashed profiles in Figure 4.4). However when we include decay (Figure 4.3), these profiles both go closer to zero, thereby reducing the difference between the two end-of-cycle dipole moments and hence increasing the *relative* dipole moment obtained by the 10 regions. This effect is even stronger for the other two cycles. Inclusion of decay does not affect the basic shape of each profile, it merely weakens the contribution from stronger regions. This can be seen by comparing the solid and dashed lines in Figure 4.4.

The $D_{\mathrm{rel}}(t_{\mathrm{end}})$ of Cycle 22 rises at a steady rate as more regions are added, but there are two clear phases with a large jump in between. One can attribute this jump to the inclusion of the largest contributor of Cycle 22. Because of this significant addition to the dipole moment, using 55% of regions is enough to ensure that 80% of $D_{\mathrm{tot}}(t_{\mathrm{end}})$ is reached.

The profile for Cycle 23 initially reaches almost $0.5\,D_{\mathrm{tot}}(t_{\mathrm{end}})$, presumably because the regions with strongest flux contribute positively to the dipole moment. There is then barely an increase in $D_{\mathrm{rel}}(t_{\mathrm{end}})$ as another 30% of the regions are included. This mimics the problem found in Figure 4.3; Cycle 23 is largely dominated by negative $D_{\mathrm{rel}}(t_{\mathrm{end}})$ active regions.

The Cycle 24 profile is on average the most effective at trying to match $D_{\mathrm{tot}}(t_{\mathrm{end}})$ when regions are ranked by flux, suggesting that Cycle 24 is dominated by large positive $D_{\mathrm{rel}}(t_{\mathrm{end}})$ regions. This is supported by Figure 4.3, which shows that when the top ten contributors are removed from the simulation, the final axial dipole moment is underestimated.

It may be noteworthy that when 60% of the strongest regions are incorporated (i.e.

regions with flux above about 2×10^{21} Mx), the four cycles all reach 80% of the final D_{tot} and adding small regions bears minimal difference, regardless of cycle number. If 90% of regions are used, corresponding to a threshold of approximately 5×10^{20} Mx, all four cycles reach a similar relative level close to $D_{\mathrm{tot}}(t_{\mathrm{end}})$.

4.3 Distributions of active region properties

4.3.1 Latitude, flux and initial dipole moment

We now turn to analyse the effects of emergence latitude, flux and initial D_{rel} on the axial dipole moment contribution $D_{\mathrm{rel}}(t_{\mathrm{end}})$ of each region. Latitude is calculated by taking the longitudinal average of the magnetogram and finding the location of the centroid of unsigned flux; applying the usual method of finding the centroid of each polarity is not necessarily meaningful when considering complex multipolar regions. For magnetic flux we integrate the radial magnetic field in latitude and longitude, and initial D_{rel} is calculated using Equation 4.1.2 at time $t = t_{\mathrm{start}}$.

The top panels of Figure 4.5 show the relationships between $D_{\mathrm{rel}}(t_{\mathrm{end}})$ and these three quantities from left to right respectively for the regions from Cycle 21. We find that most significant contributors to the axial dipole moment emerge below $\pm 20°$, the very largest of which emerge below $\pm 10°$. We also find that these regions do not necessarily have strong levels of magnetic flux; very few of the biggest contributors are stronger than 1.5×10^{22} Mx.

We see that the relationship between initial and final D_{rel} is largely determined by the emergence latitude: regions emerging at mid-latitudes (dark purple) tend to contribute little to the final axial dipole moment, regardless of their initial values. Conversely, regions emerging at low latitudes (yellow and orange) can undergo an increase in axial dipole moment contribution as cross-equatorial flux cancellation occurs and flux is transported poleward by the meridional flow.

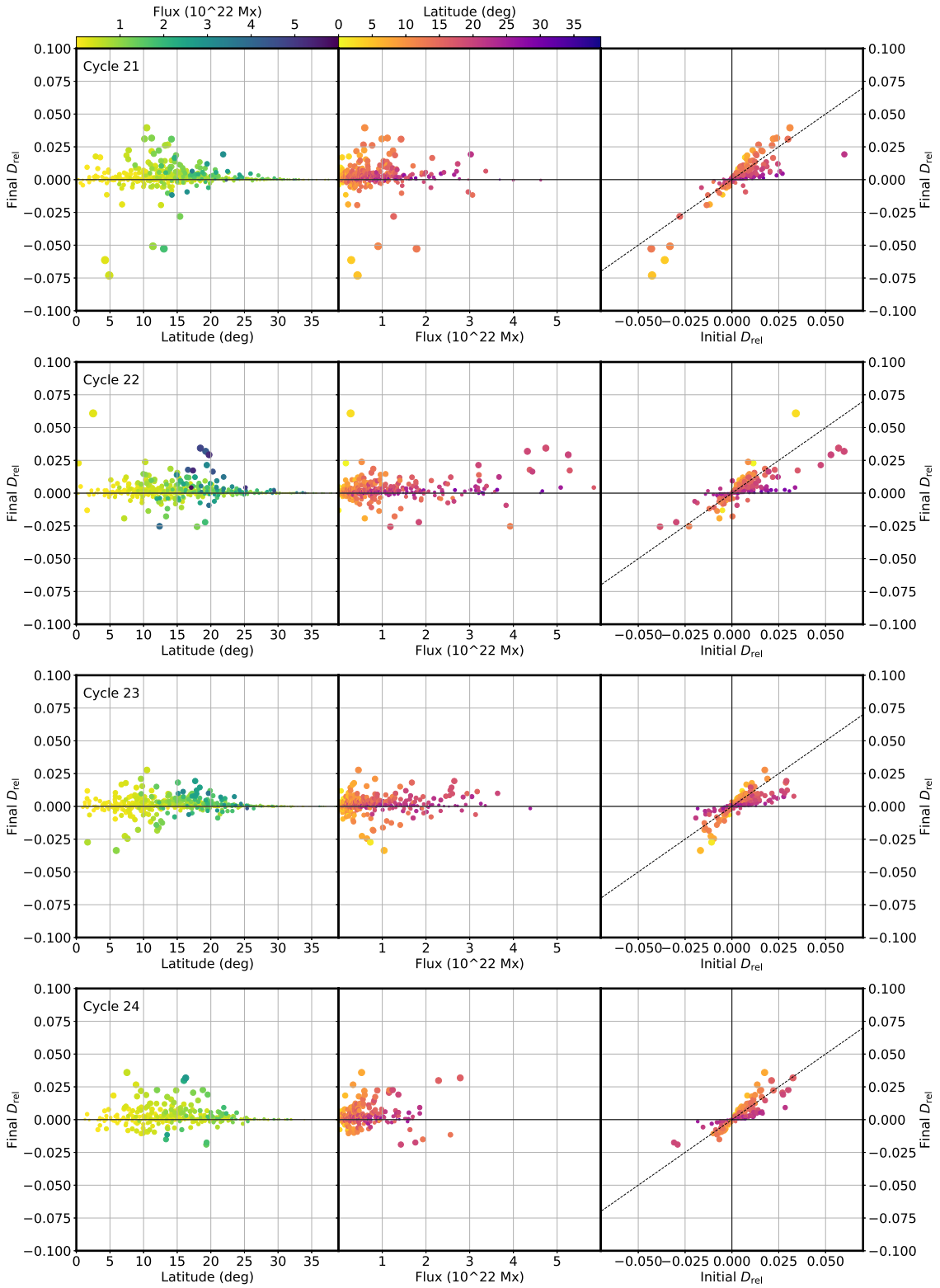

Figure 4.5: Final D_{rel} for each region against absolute latitude (left panels), flux (middle panels) and initial D_{rel} (right panels). Markers are sized by absolute final D_{rel}, and coloured by flux (left panels) and absolute latitude (middle and right panels).

The second row of Figure 4.5 shows the same relationships as discussed above but for Cycle 22. The left and middle panels tell a different story to that of Cycle 21. There are fewer big contributions (i.e. contributions of more than 2.5%) to the axial dipole moment, and the largest is a strengthening rather than a weakening as in Cycle 21. This explains why the axial dipole moment increased in amplitude during Cycle 22, and why polar field reversal is almost achieved with just ten regions in Figure 4.3(a). This largest region is also the only significant contributor to lie below $\pm 10°$, although the others still emerge below $\pm 20°$ as in Cycle 21. The most striking difference between the two cycles is the effect of strong-flux regions. In Cycle 22 some of the most significant contributions to the axial dipole moment come from regions with fluxes above 3×10^{22} Mx, which is not the case in Cycle 21. The same latitudinal dependence of the initial to final D_{rel} relationship is found as in Cycle 21, supporting the idea that latitude of emergence plays an important role in determining whether a region will contribute significantly to the polar field.

The third row of panels in Figure 4.5 shows the same three distributions but for Cycle 23. We return to a similar regime to Cycle 21: of the most significant contributors, we observe more regions which weaken the axial dipole moment, and the biggest contributors have fluxes smaller than 2×10^{22} Mx. Again, most of these regions emerge below $\pm 20°$. We find that the most significant regions in Cycle 23 induce a weakening of the overall axial dipole moment. These low-latitude regions could indeed be the cause of the weak polar field at the end of Cycle 23, and hence the low amplitude of Cycle 24, as suggested by Jiang *et al.* (2015).

The final row of panels shows the distributions for Cycle 24. We see that regions that emerge above $\pm 20°$ are not significant contributors to the dipole moment. Again, we see a similarity between Cycles 22 and 24, in that both big and small regions (in terms of flux) can be large contributors. However, we should eventually also consider active regions from the remainder of the cycle, as there is a small chance that regions emerging later in the cycle will significantly add to these statistics.

The latitude-dependent relationship between initial and final D_{rel} holds across all

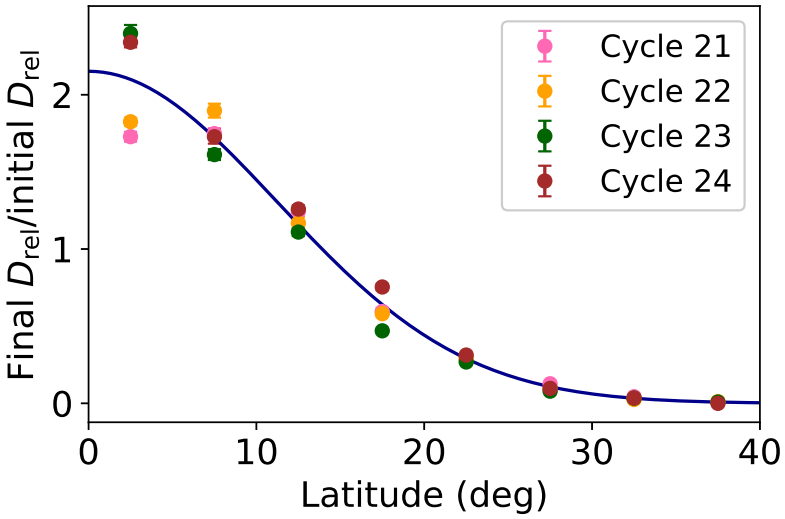

Figure 4.6: Ratio between final D_{rel} and initial D_{rel} for 5° latitudinal bins for Cycles 21 (pink), 22 (yellow), 23 (dark green) and 24 (brown). Error bars show standard error. Markers are plotted at the midpoint of each 5° bin. The dark blue curve is a Gaussian fit to the data.

cycles. Separating the regions into bins of 5° and calculating the gradient of the lines in the right-hand panels of Figure 4.5 for each bin (see Figure 4.6), we find that down to ±20° the relationship between initial and final D_{rel} is practically identical across the four cycles, and even down to ±5° the relationships over the four cycles are close. For the 0–5° bin, the gradient is much steeper for Cycles 23 and 24. However, this bin has relatively few points, and is least well fitted by a linear relationship between initial and final D_{rel}. The standard errors for all of the fits are very small, indicating a strong relationship between the overall amplification in D_{rel} and the latitude of emergence. If we fit a Gaussian to the data (dark blue curve in Figure 4.6), we find that the axial dipole moment contribution is proportional to $\exp\left(-\frac{\lambda^2}{252}\right)$. This is similar to the relationship between latitude and axial dipole moment contribution given by Jiang *et al.* (2014) who also found a Gaussian latitudinal dependence in their model (Equation 4.0.1). The difference in Gaussian width is probably caused by differences in parameter choices, particularly meridional flow velocity.

4.3.2 Latitude and time

We now focus on the time-latitude distributions, i.e. 'butterfly diagrams', of the active regions drawn from the assimilative 2D model. The top section of Figure 4.7 shows the butterfly diagrams of Cycle 21 for the cases shown in the first section of Figure 4.3(a), where border colours match profile colours. We find few strong regions that have emerged across the equator, suggesting that large contributors from Cycle 21 are likely to be because of orientation reasons rather than being cross-equatorial. There is a cluster of negatively contributing regions in the northern hemisphere around 1983 which is not followed by many significant regions during the remainder of the cycle; this cluster could be responsible for a lower axial dipole moment in Cycle 21 (compared to Cycle 22), and explains why the polar field fails to reverse when only 10 regions are used in Cycle 21, as seen in Figure 4.3(a).

The bottom section of Figure 4.7 shows the corresponding butterfly diagrams for Cycle 22. As inferred from Figure 4.5, the majority of large contributions to the axial dipole moment in Cycle 22 enhance the dipole moment and are clustered around $-20°$. However, there are two large contributors at low latitudes, possibly cross-equatorial, which would support the claim of Cameron *et al.* (2013): that regions emerging across the equator can significantly change the amount of net flux in each hemisphere, in turn weakening or strengthening the axial dipole moment, meaning future cycle predictions will be less reliable.

The top section of Figure 4.8 shows the butterfly diagrams of Cycle 23. Significant negatively-contributing regions include a cluster across the equator around 2002, and a group of regions in the southern hemisphere towards the end of the cycle, visible as blue patches in all but the bottom-right frame. While the cross-equatorial group is important for reasons discussed above, the majority of regions in the late-emerging cluster might not have had as significant an effect on the current cycle as if they had instead emerged earlier in the cycle, as discussed by Nagy *et al.* (2017), who inserted an extreme active region into a dynamo model simulation at different times

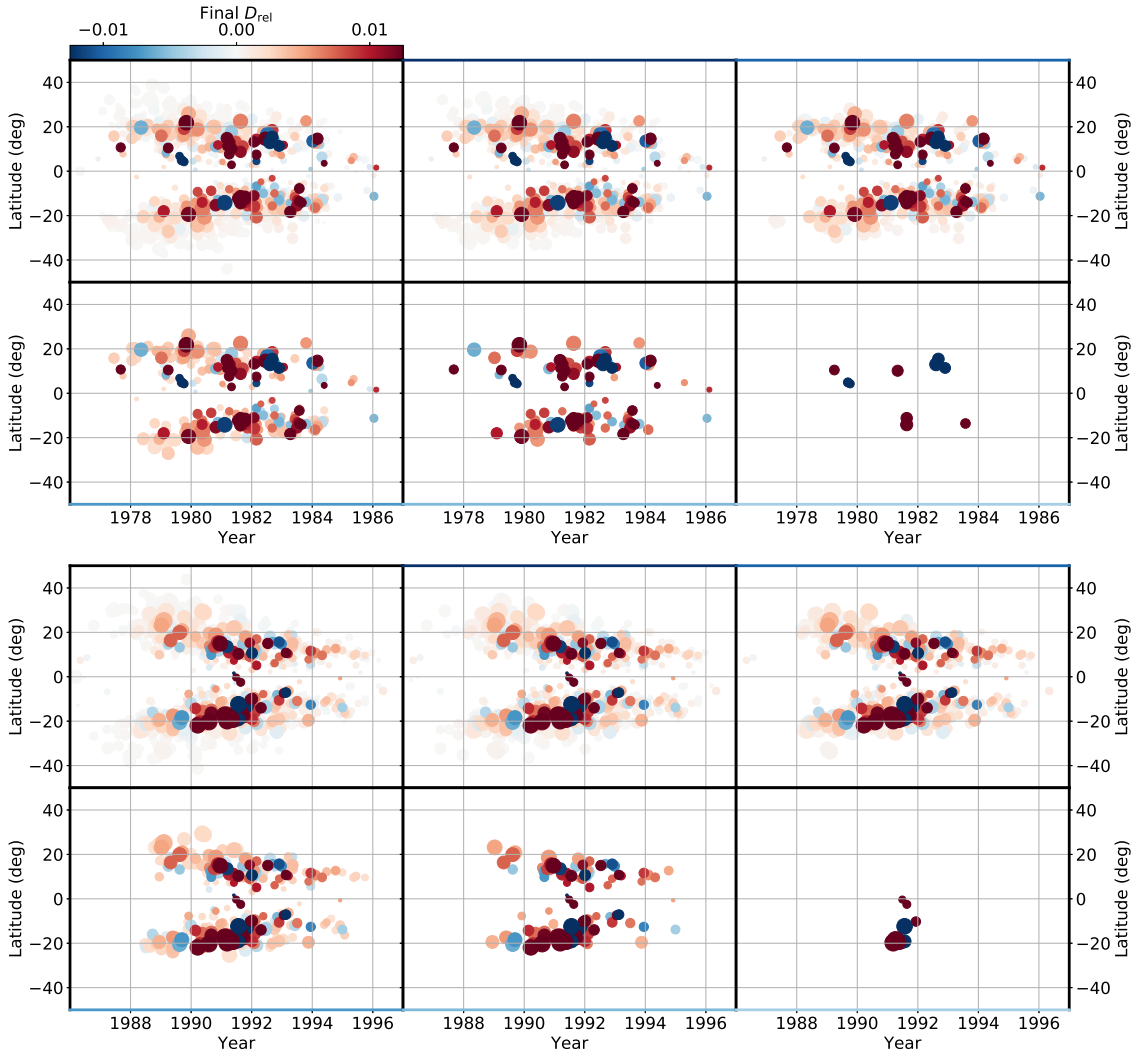

Figure 4.7: Time-latitude distributions of regions from Cycles 21 (top) and 22 (bottom) used to obtain the profiles in the first and second sections of Figure 4.3(a), where border colours match profile colours. That is, the bottom right panel shows the top 10 regions, increasing from right to left, bottom to top. Markers are sized by flux and coloured by final D_{rel}.

throughout a cycle and found that late-emerging regions had the smallest effect. This is because any poleward-advected flux would not have had enough time to reach the pole and cancel with the polar field before the end of the cycle. The weaker contribution from regions emerging later in the cycle is also evident in Figure 4.7, suggesting that it could take at least a few years for regions to reach their asymptotic contributions to the axial dipole moment. Nevertheless, by comparing Cycles 21 and 23 we see that a lack of disruption from a major cross-equatorial region in Cycle 21 led to a stronger axial dipole moment compared to Cycle 23. The butterfly diagrams again illustrate that the largest contributors are not necessarily the biggest in terms of flux.

Finally, the bottom section of Figure 4.8 shows the equivalent butterfly diagrams for Cycle 24. Uniquely, we find that the top 10 most significant contributors are all positive, indicating that Cycle 24 has behaved in an 'ordinary' way and has not significantly altered the amplitude of the next cycle. This is supported by the precursor predictions discussed in Section 1.2.2, which generally state that Cycle 25 is likely to be of similar strength to Cycle 24.

4.4 Effect of decay on the axial dipole moment

As mentioned in Section 4.1, we also remove the decay term from Equation 2.1.1 (i.e. set $\tau \to \infty$) and repeat the optimization and subsequent analysis on the same four cycles. Whilst the equivalent distributions as those shown in the scatterplots of Section 4.3.1 and butterfly diagrams of Section 4.3.2 are qualitatively indistinguishable up to a scaling factor, the axial dipole moment profiles for simulations with regions included based on $D_{\mathrm{rel}}(T)$ as shown in Section 4.2 behave slightly differently, simply because of the lack of decay impacting on cycle minima.

The profiles from simulations without decay where only the largest contributors are included are shown in Figure 4.9(a). With less freedom from fewer parameters, the optimal axial dipole moment does not match the observed counterpart as well

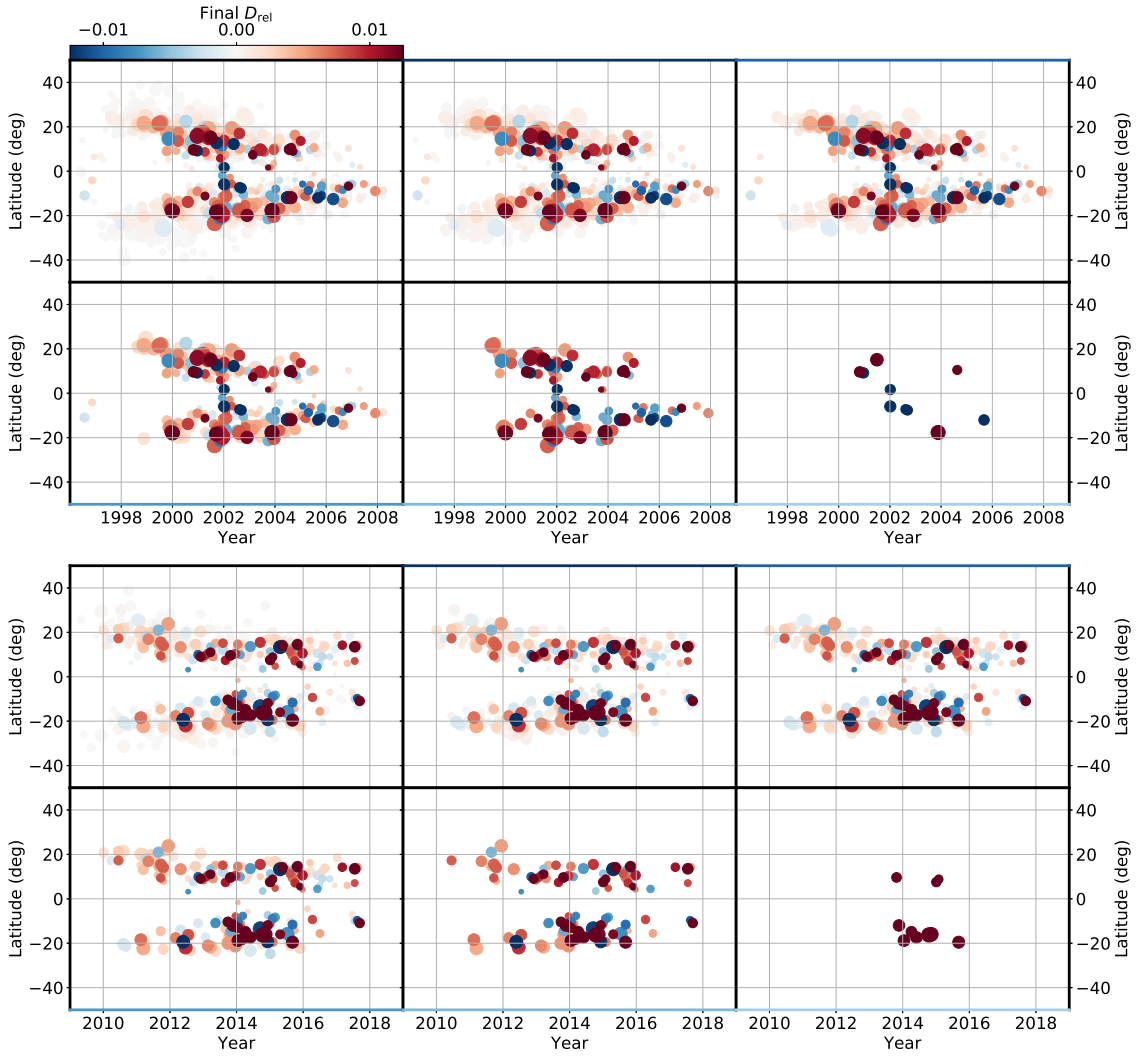

Figure 4.8: Time-latitude distributions of regions from Cycles 23 (top) and 24 (bottom) used to obtain the profiles in the third and fourth sections of Figure 4.3(a), where border colours match profile colours. That is, the bottom right panel shows the top 10 regions, increasing from right to left, bottom to top. Markers are sized by flux and coloured by final D_{rel}.

when decay is included, but the fit is still acceptable. Again we find that when the top 750 (400) contributors are used, Cycles 21, 22 and 24 are hardly affected but the discrepancy in Cycle 23 is now even more visible than before. When the 100 largest contributors are used, the polar field reverses in Cycles 21, 22 and 24, but not in Cycle 23. Furthermore, polar field reversal is only just achieved with 250 regions in Cycle 23, supporting the claim that the biggest contributors from Cycle 23 contribute negatively to the axial dipole moment. With their BMR data for Cycle 21, Wang & Sheeley (1991) found that about 54% of the axial dipole moment came from about 10.7% of regions, and here we find a similar result (dashed blue curve). In fact, we find the same outcome for Cycle 22 but not for Cycle 23.

Figure 4.9(b) shows the axial dipole moment evolution when the strongest regions are removed from each cycle. With no exponential decay, the deficit created by the removal of the top 10 regions of Cycle 23 is even clearer here than in Figure 4.3(b), highlighting the detrimental effect of those contributors with negative $D_{\mathrm{rel}}\,(T)$.

4.5 Conclusions

Our aim was to test claims that the polar field at the end of Cycle 23 could have been weakened by a small number of large, low-latitude regions. We extracted active region properties from magnetograms using an automated region assimilation technique, and analysed the relationships between these properties and the evolution of the axial dipole moment using a 2D flux transport model.

We first looked at the effect of keeping regions with the largest final axial dipole moment contribution $D_{\mathrm{rel}}\,(t_{\mathrm{end}})$ in the simulation in increments, to see how many were required to obtain a good match with the original axial dipole moment. Using the 500 (or 300 for Cycle 24) biggest contributors produced an acceptable axial dipole moment in Cycles 21, 22 and 24, but the lack of small contributions was more damaging in Cycle 23, where at least 750 regions are required to produce an acceptable match. When we only considered the top 10–100 regions, we observed

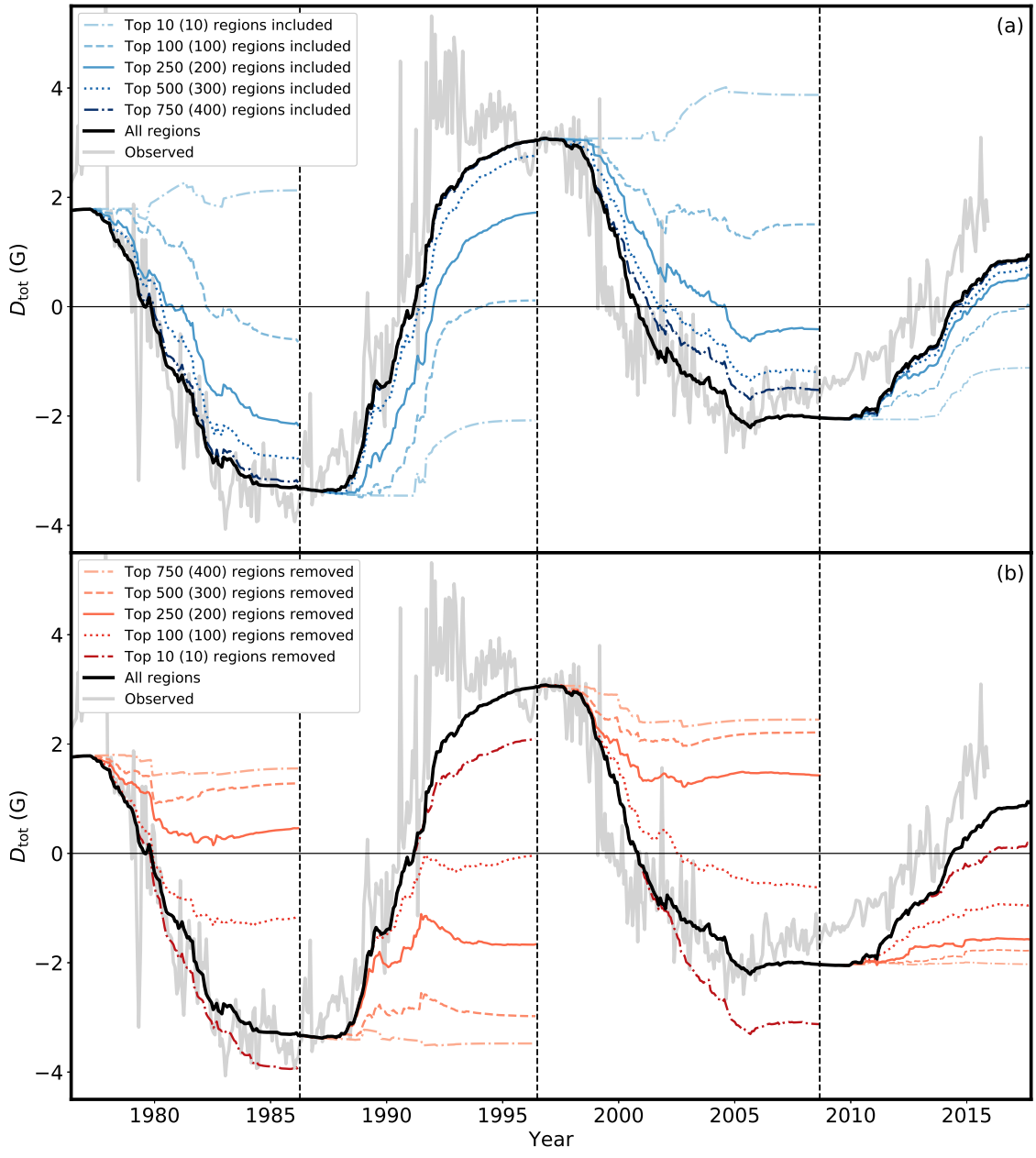

Figure 4.9: Evolution of the axial dipole moment for Cycles 21 to 24, equivalent to Figure 4.3 but with no exponential decay term.

that the odd-numbered cycles, especially Cycle 23, struggled to achieve polar field reversal. We attributed this discrepancy to the influence of negatively contributing regions which appear to dominate the axial dipole moment. On the removal of these strongest contributors we found that the axial dipole moment was enhanced, suggesting that the weak polar field at the Cycle 23/24 minimum may indeed have been caused by a small number of extreme regions. When regions were included in order of flux instead of $D_{\mathrm{rel}}\,(t_{\mathrm{end}})$ there were some differences between cycles, although in each case using the top 80–90% of the strongest regions was enough to provide a good match to the original axial dipole moment.

We also examined how the final contribution of a single region to the axial dipole moment at the end of the cycle is affected by a region's emergence latitude, flux and initial axial dipole moment, and compared these relationships across Cycles 21 to 24. We found that generally all large contributions to the axial dipole moment emerge below $\pm 20°$, with the largest emerging below $\pm 10°$. This supports the idea that regions emerging at low-latitude can have a large effect on the evolution of the axial dipole moment (Cameron *et al.*, 2013; Jiang *et al.*, 2015). For our more realistically shaped multipolar regions, we cannot measure the conventional tilt angle, so instead we calculated the more meaningful parameter of initial relative axial dipole moment which takes into account orientation as well as latitude. We found a positive correlation between initial and final D_{rel} within all latitudinal bins in all cycles, but that the constant of proportionality depended on latitude with regions at low latitudes contributing most, whence we concluded that emergence latitude is the dominant parameter controlling the amplification or suppression of the initial dipole moment of a region. This latitude dependence exists because a large dipole moment arises from hemispherical polarity separation, which occurs most effectively when regions emerge tilted and at low latitudes so that cross-equatorial transport of flux can occur (Wang & Sheeley, 1991; Yeates *et al.*, 2015). Therefore once we have measured the initial dipole moment of a given region, we can predict its long-term contribution to the dipole moment based purely on its latitude of emergence and

the flux decay parameter τ.

We found that the patterns of regions contributing most to the dipole moment were not consistent across the three cycles. In particular, Cycle 22 contained multiple strong-flux regions which were also some of the largest contributors to the axial dipole moment. This was not the case in Cycles 21 and 23; most large contributors had fluxes of less than 2×10^{22} Mx, reinforcing that flux alone is not an appropriate measure of contribution. Incidentally, across all cycles there were no significant contributors with fluxes less than 1×10^{21} Mx, indicating that the smallest regions are not able to drastically alter the axial dipole moment, regardless of emergence latitude. In their coupled surface-interior model, Nagy *et al.* (2017) showed that changing BMR tilt and emergence latitude had more immediate consequences than changing flux, unless a very large amount of flux was included. Consequently, if a very large, anti-Joy, anti-Hale region was to emerge close to the equator, it could have a significant detrimental impact on the polar field and hence the amplitude of the next cycle. Following the results of Nagy *et al.* (2017) it could even be speculated that, in the most extreme case, such an event could lead to a grand minimum.

As discussed in Chapter 1, some predictions of Cycle 25 using the polar field as a precursor have already been made, for example by Hathaway & Upton (2016) and Cameron *et al.* (2016), who used two distinct models but came to a similar conclusion: that Cycle 25 will be another weak cycle. However, by incorporating uncertainty in tilt angles and performing multiple simulations, a wider range of cycle amplitudes was found, suggesting that the behaviour of our Sun really does hinge on the random fluctuations in active region properties, highlighting the incurred uncertainty in making *early* forecasts of the next cycle, and that making predictions of future cycles is perhaps futile.

On a more positive note, as we approach the minimum at the end of Cycle 24 predictions of Cycle 25 will become more reliable, since it becomes less likely that any more large regions which can significantly alter the polar field will emerge. Indeed, from our analysis of the previous three cycles, we only found significant

contributors emerging up to the early stages of the descending phase, although that isn't to say such an event is not possible. Indeed, Nagy *et al.* (2017) found that 'rogue' regions emerging late in the cycle can still have an effect on the following cycle, but this cannot be assessed using our surface flux transport approach, and requires simulation of the interior of the convection zone. For completeness we should go back and repeat this analysis once we reach cycle minimum, using the results to assess any current predictions of Cycle 25.

Chapter 5

Three-dimensional kinematic dynamo model

Progression in Babcock-Leighton models of the solar dynamo has primarily been through the implementation of 2D or 2×2D models (e.g. Wang *et al.*, 1991; Durney, 1995; Chatterjee *et al.*, 2004; Guerrero & de Gouveia Dal Pino, 2008; Lemerle & Charbonneau, 2017). However, we would ideally like to develop 3D B-L dynamo models in order to realistically model the emergence of buoyant magnetic structures, and fully describe the evolution of magnetic fields under the effects of diffusion, differential rotation and meridional circulation. These models are more complex and require in-depth calibration in order to match the observed magnetic field. Nevertheless, success in overcoming these obstacles would be a sizeable step towards the development of a forecasting model for the Sun-Earth system (Nita *et al.*, 2018), and would hopefully provide us with the most accurate solar cycle predictions to date.

Yeates & Muñoz-Jaramillo (2013) developed KD3, a 3D kinematic B-L dynamo model which was the first of its kind, though some authors have since had success in developing their own 3D B-L dynamo models (Miesch & Dikpati, 2014; Kumar *et al.*, 2018). In KD3, the ideal MHD induction equation describes the evolution of

the magnetic field:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) - \nabla \times (\eta \nabla \times \mathbf{B}), \tag{5.0.1}$$

for a prescribed velocity field $\mathbf{u}(r, \theta, \phi, t)$ and a prescribed turbulent diffusivity $\eta(r)$. There is no small-scale α-effect. Equation 5.0.1 is solved in a spherical shell using a finite volume scheme. The coordinates are not actually spherical but stretched Cartesian (van Ballegooijen *et al.*, 2000), with variable grid spacing in the horizontal directions. For more details see Appendix A of Yeates & Muñoz-Jaramillo (2013).

Unlike previous 2D B-L dynamo models, KD3 explicitly models the buoyant emergence of flux tubes through the convection zone (Fan, 2009). In the 2D models, the active region emergence process has either been parametrized through a volumetric α-effect term in the induction equation, or through manual insertion of regions at the surface, corresponding to areas of strong toroidal field at the base of the convection zone (e.g. Durney, 1997; Nandy & Choudhuri, 2001; Muñoz-Jaramillo *et al.*, 2010; Guerrero *et al.*, 2012). However, these 'non-local' methods make magnetic flux conservation difficult to enforce because the process of forming the emerging region from the pre-existing toroidal field is not followed explicitly through the induction equation (5.0.1). In KD3, a time-dependent velocity perturbation $\mathbf{v}$ is included which is intended to capture the effects of advection and buoyancy on the flux tubes. The non-axisymmetric perturbation has a radial component, which transports the tube outwards through the convection zone to the surface; a vortical component, which models the helical convective motions and gives rise to tilts in the active regions; and a diverging component, responsible for expanding the tube as the density decreases. The tube centre velocity is set so that the travel time from $r = 0.7\,R_\odot$ to $r = R_\odot$ is 25 days, after which the perturbation is removed.

Yeates & Muñoz-Jaramillo (2013) presented a full simulation of Solar Cycle 23, using BMR data from NSO/Kitt Peak, and showed that the KD3 model is able to reproduce the qualitative behaviour of active region decay at the surface, leading to poleward transport of flux and reversal of the polar field. However, the photospheric magnetic

field was not systematically calibrated against observed data, and closer inspection has shown that the quantitative details of the surface evolution are significantly different from 2D SFT models, even when the same parameters are used at the surface. Moreover, Yeates & Muñoz-Jaramillo (2013) did not demonstrate a self-sustaining dynamo. A future goal is to use the genetic algorithm `PIKAIA` from Chapter 2 to optimize parameters for the 3D model to produce the most accurate dynamo simulations to date, and eventually develop a sun-like, self-excited dynamo model. This would be a huge step forward in solar cycle modelling and prediction.

In this chapter we explore the effects of radial diffusion in KD3 which we believe to be the cause of the discrepancy between the SFT model and dynamo model. In Section 5.1 we present the inconsistency, and in Section 5.2 we investigate the effect of radial diffusion using a simple Cartesian 2D model of the convection zone. We return to KD3 in Section 5.3 to assess whether the same results hold. In Section 5.4 we attempt to simulate Cycle 23 using an 'improved' diffusion profile, and in Section 5.5 we conclude. For the published version of this chapter, see Whitbread *et al.* (2019).

5.1 Radial diffusion in KD3

On comparing the surface evolution of B_r in KD3 with the 2D SFT model from earlier chapters, it is apparent that the two are markedly different, even when the same horizontal flows and diffusivity and same initial B_r are used at the surface. As an example, the SFT evolution of a single BMR (Figure 5.1) placed at 10° latitude with flux 1×10^{22} Mx and a tilt angle of 30° is shown in the top panel of Figure 5.2, and the KD3 equivalent is shown below. The parameters used are introduced in Section 5.3. The BMR is inserted in the SFT simulation at the time when the flux has stopped emerging in KD3, i.e. when the unsigned flux at the photosphere has reached its peak (Figure 5.3). Even though the differential rotation, meridional flow and horizontal diffusion in the SFT model match the surface parameters of the

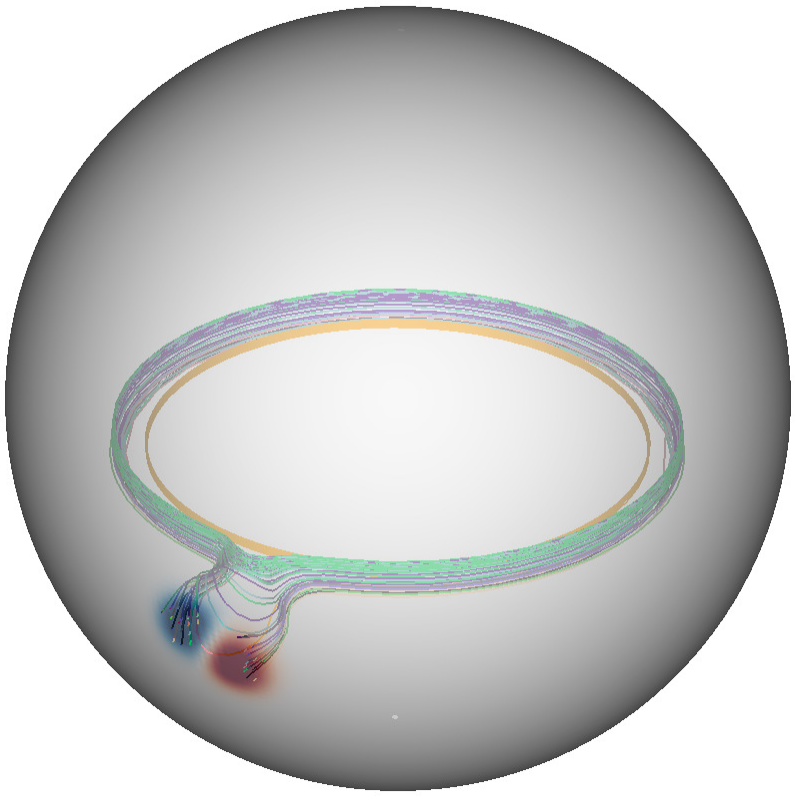

Figure 5.1: Three-dimensional image of an emerged active region in KD3. Magnetic field lines are connected to the toroidal field at the base of the convection zone and the radial magnetic field is shown at the transparent surface.

KD3 simulation, the transport to the poles is significantly faster in the SFT case. In addition, the top panel of Figure 5.3 shows that there is significantly more flux present at the surface in the KD3 system. There is also a large difference in the respective evolutions of the polar flux (bottom panel of Figure 5.3). In KD3, the south polar field barely develops by the end of the simulation, and the peak of the north polar field is stronger and occurs 3 years later than in the SFT case.

In Section 1.3.2 we took the radial component of the induction equation (Equation 5.0.1) and neglected radial derivatives to derive the surface flux transport model. If we instead include radial derivatives, the advection term remains the same but the

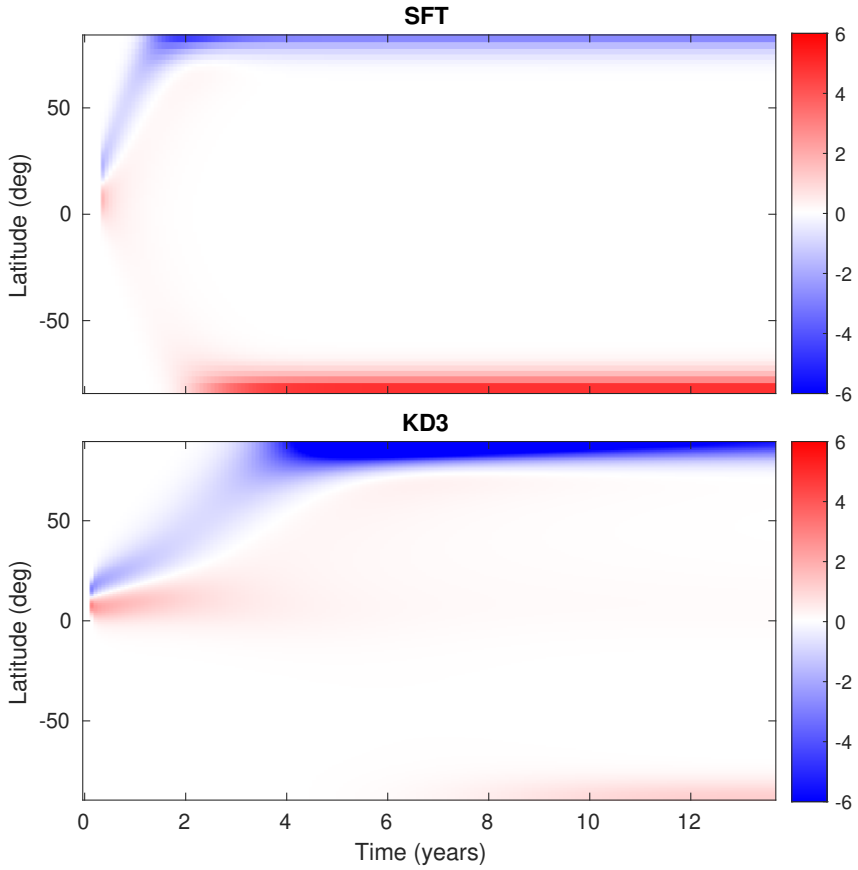

Figure 5.2: Top: Longitude-averaged evolution of B_r for a single BMR in a 2D SFT model. Bottom: Surface component of the 3D dynamo model showing the equivalent evolution of the same BMR.

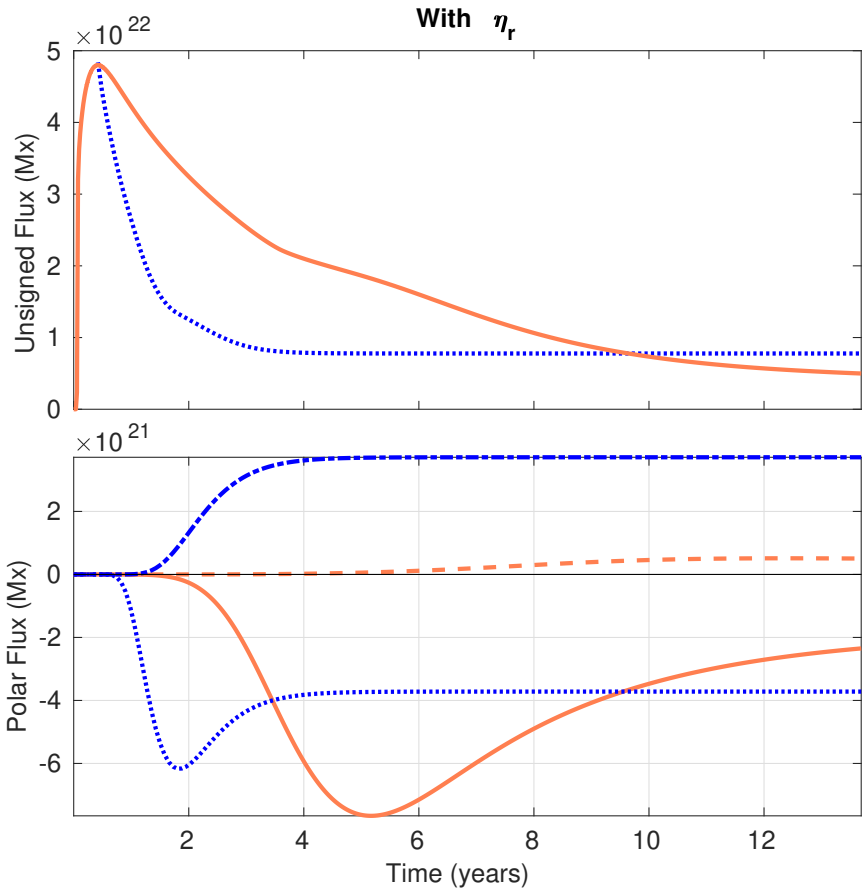

Figure 5.3: Top: Comparison of unsigned surface flux from the 2D SFT simulation (blue) and 3D dynamo simulation (orange). Bottom: Comparison of northern (solid and dotted lines) and southern (dashed and dash-dotted) polar flux from the same two simulations, where polar flux is defined as the flux poleward of 70° latitude.

diffusion term gains extra components:

$$
\begin{aligned}
[\nabla \times (\eta \nabla \times \mathbf{B})] \cdot \mathbf{e_r} = & \Bigg[\nabla \times \Bigg\{ \left(\frac{\eta}{R_\odot \sin\theta} \frac{\partial B_r}{\partial \phi} - \frac{\eta}{r} \frac{\partial}{\partial r} (r B_\phi) \right) \mathbf{e}_\theta \\
& + \left(\frac{\eta}{r} \frac{\partial}{\partial r} (r B_\theta) - \frac{\eta}{R_\odot} \frac{\partial B_r}{\partial \theta} \right) \mathbf{e}_\phi \Bigg\} \Bigg] \cdot \mathbf{e_r} \\
= & - \frac{\eta}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial B_r}{\partial \theta} \right) - \frac{\eta}{r^2 \sin^2\theta} \frac{\partial^2 B_r}{\partial \phi^2} \\
& + \frac{\eta}{r \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial B_\theta}{\partial r} \right) + \frac{\eta}{r \sin\theta} \frac{\partial^2 B_\phi}{\partial \phi \partial r}. \qquad (5.1.1)
\end{aligned}
$$

The final two terms are not present in the SFT equation, in the absence of radial derivatives. It is likely that the discrepancy shown in Figs. 5.2 and 5.3 is due to this exclusion of radial diffusion in the SFT model. As discussed in earlier chapters, some attempts to mimic the effect of radial decay have been implemented. For example, an exponential decay term of the form $-\frac{B_r}{\tau}$ was added to the model by Schrijver *et al.* (2002) to ensure regular polar field reversal over many cycles, and Baumann *et al.* (2006) gave a physical explanation of this decay term: it is the effect of radial diffusion of flux into the solar interior, which cannot be accounted for directly in the SFT model. Baumann *et al.* (2006) built on this work by considering a modal version of the decay term, where a poloidal magnetic field in the convection zone is decomposed into spherical harmonic form and the decay time of each mode is calculated. This is in contrast to the exponential decay parameter used by Schrijver *et al.* (2002), where all modes decay at the same rate. Despite this added layer of sophistication, it is not trivial to say whether this simplified representation of radial diffusion is realistic. Whitbread *et al.* (2017) and Virtanen *et al.* (2017) have shown that using modal decay rates does not make a significant difference to the butterfly diagram.

It should be noted that, like in previous chapters, the diffusivity η is merely a turbulent diffusivity, representing the diffusive effect of convection. Horizontally, this corresponds to random walks over the supergranular convection cells, and vertically it mimics the effect of opposing motions of upward convective plumes and down-

ward pumping. The buoyancy in the model is handled by the emergence velocity perturbation. Ideally we would like to include convective motions in the model as advection rather than diffusion, but it is currently too expensive computationally in 3D, although see Upton & Hathaway (2014*b*) for a purely advective 2D SFT model.

Using $\nabla \cdot \mathbf{B} = 0$ in spherical coordinates, we have:

$$-\frac{\sin\theta}{r}\frac{\partial}{\partial r}\left(r^2 B_r\right) = \frac{\partial}{\partial\theta}\left(\sin\theta B_\theta\right) + \frac{\partial B_\phi}{\partial\phi}, \tag{5.1.2}$$

so the radial component of the diffusion term from the induction equation can be rewritten as:

$$\begin{aligned}[\nabla \times (\eta\nabla \times \mathbf{B})] \cdot \mathbf{e_r} = &-\frac{\eta}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial B_r}{\partial\theta}\right) - \frac{\eta}{r^2\sin^2\theta}\frac{\partial^2 B_r}{\partial\phi^2} \\ &-\frac{\eta}{r}\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r^2 B_r\right)\right).\end{aligned} \tag{5.1.3}$$

If we remove the final term from the right-hand side of Equation 5.1.3 from the KD3 model, we find that the model now behaves similarly to the SFT equivalent, as shown in Figure 5.4. Again, the surface parameters are identical and we have inserted the BMR in the SFT model once all flux has emerged in KD3. The unsigned flux is shown in the top panel of Figure 5.5, and this confirms the reconciliation between the two models. Furthermore, in the bottom panel of the same figure we see that the qualitative evolution of the polar flux in each model is the same. There is a small discrepancy in both panels due to subtle differences between the models, such as the numerical grid, but the general evolution is consistent.

However, this removal of the radial diffusion term in KD3 is artificial, and the two models should reproduce at least similar results at the surface with all terms included. Moreover, we have shown in Chapter 2 that the SFT model is able to sufficiently match the observed magnetic field, so the likelihood is that the problem lies with the KD3 model and we need to fix it, or at the very least improve it, without removing the radial diffusion term. This should be done before any attempts to optimize parameters or search for self-sufficient dynamo solutions are made.

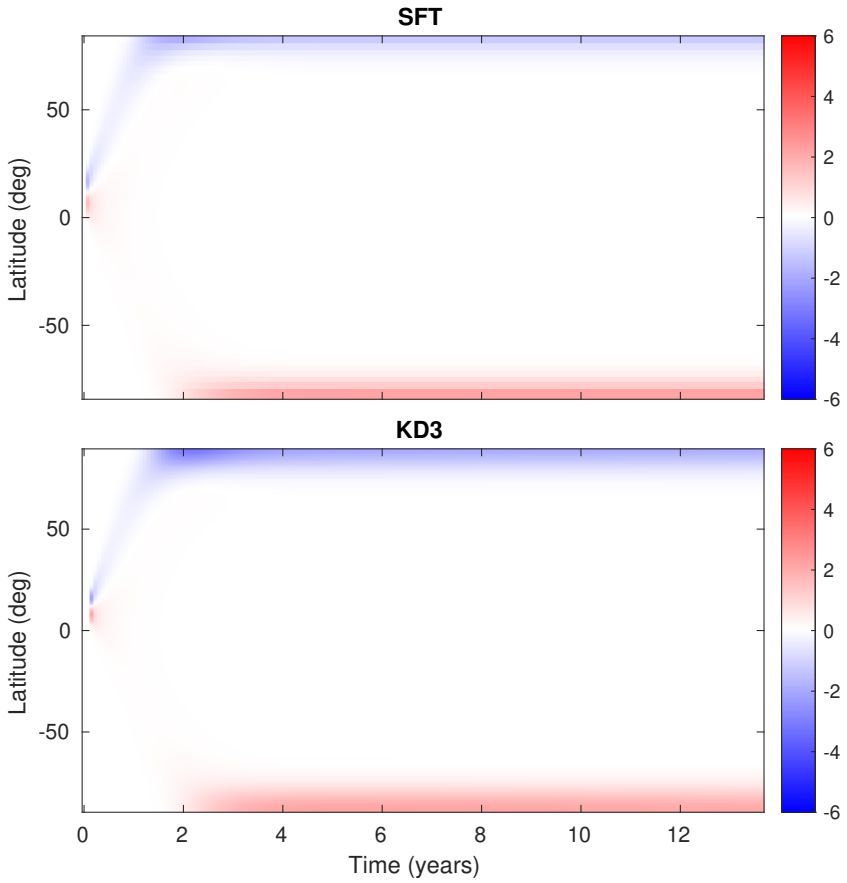

Figure 5.4: Top: Butterfly diagram showing the 2D surface evolution of B_r for a single BMR. Bottom: Surface component of the 3D dynamo model showing the equivalent evolution of the same BMR but with the radial diffusion term removed.

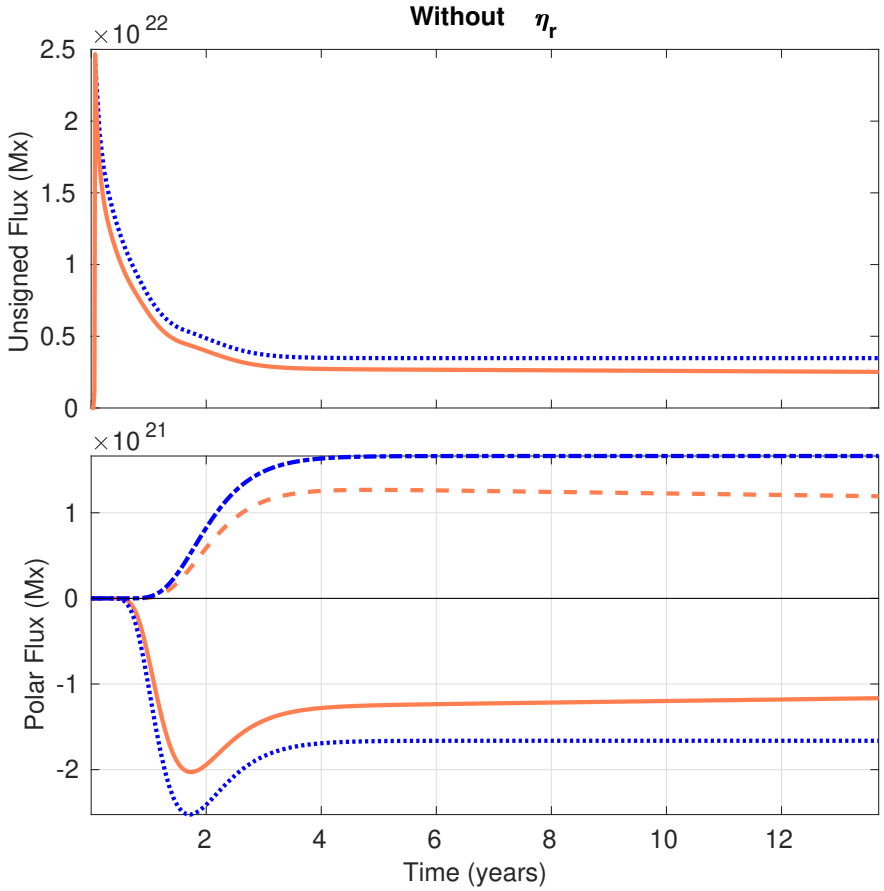

Figure 5.5: Top: Comparison of unsigned surface flux from the 2D SFT simulation (blue) and 3D dynamo simulation with the radial diffusion term removed (orange). Bottom: Comparison of northern (solid and dotted lines) and southern (dashed and dash-dotted) polar flux from the same two simulations, where polar flux is defined as the flux poleward of 70° latitude.

5.2 2D model of active region decay

We begin by investigating a 2D model that illustrates the basic cause of the difference between the KD3 and SFT models. Inspired by van Ballegooijen (1998), we take a 2D Ω-loop representing a newly-emerged BMR in the convection zone and evolve it according to diffusion alone. The benefit of a simpler toy model is that it captures the diffusive effects of a 3D model but is computationally less expensive, and at this stage we are not interested in other features such as the amount of poloidal field produced.

Here we use Cartesian coordinates (x,z) which denote the width and depth of the convection zone domain respectively, with $-0.4 \leq x \leq 0.4$ and $0.6 \leq z \leq 1$. Neglecting variation in the y-direction, we write $\mathbf{B}$ in terms of a flux function as:

$$B_x = -\frac{\partial A}{\partial z}, \tag{5.2.1}$$

$$B_z = \frac{\partial A}{\partial x}. \tag{5.2.2}$$

We neglect the effects of advection in Equation 1.3.16, which therefore reduces to

$$\frac{\partial A}{\partial t} = \eta\,(z)\,\nabla^2 A. \tag{5.2.3}$$

Note that we allow the diffusivity η to be a function of z, so that we can investigate the effect of different diffusivity profiles with depth. The effect of advection will be considered in the 3D simulations of Section 5.3. We also simultaneously evolve a 1D surface diffusion model as the analogue of the SFT model. For visualization a potential field extrapolation is performed in the corona. As an initial condition, the region is assumed to have emerged and is connected to the toroidal field at the base of the convection zone (see top left panel of Figure 5.7), as in KD3, and is of the form:

$$A_0 = \exp\left(-\frac{z-0.6}{0.04}\right) + \frac{1}{2}\exp\left(\frac{(z-1)^2}{0.4} - \frac{x^2}{0.008}\right). \tag{5.2.4}$$

We impose periodic boundary conditions in x and set $\dfrac{\partial A}{\partial t} = 0$ at the base ($z = 0.6$).

At the surface ($z = 1$) we follow van Ballegooijen (1998) and van Ballegooijen & Mackay (2007) by setting:

$$B_{x,\mathrm{cz}} = \beta B_{x,\mathrm{cor}}, \tag{5.2.5}$$

where $B_{x,\mathrm{cz}}$ is the horizontal field at the convection zone boundary and $B_{x,\mathrm{cor}}$ is the horizontal component of a potential extrapolation into the corona. Then the parameter β determines whether the interior field at the photosphere is matched to the potential field in the corona ($\beta = 1$), or whether it is purely radial ($\beta = 0$), which was the original boundary condition in KD3 (and, indeed, in most other models). This will allow us to assess the effect of the top boundary condition on radial diffusion, although for most tests we set $\beta = 0$. The effects of changing the top boundary condition are discussed later.

In general we will use the following depth-dependent two-step profile for $\eta(z)$:

$$\eta(z) = \eta_c + \frac{\eta_0 - \eta_c}{2}\left(1 + \mathrm{erf}\left(\frac{z - R_1}{\Delta_1}\right)\right) + \frac{\eta_s - \eta_0 - \eta_c}{2}\left(1 + \mathrm{erf}\left(\frac{z - R_2}{\Delta_2}\right)\right), \tag{5.2.6}$$

which is normalised so that $\eta_{\max} = 1$. Here η_c is the core diffusivity, η_0 is the diffusivity in the convection zone, and η_s is the surface diffusivity. The step locations and thicknesses are R_i and Δ_i respectively. The profiles used in this thesis are shown in Figure 5.6. For a given diffusion profile and boundary condition, Equation 5.2.3 is solved using an explicit second-order finite difference method with Euler timestepping.

For comparing different diffusivity profiles, it is useful to calculate the apparent velocity of the field lines (Wilmot-Smith *et al.*, 2005), using the fact that the contours of A are the field lines of $\mathbf{B}$, since $\mathbf{B} \cdot \nabla A = 0$. We have the equation:

$$\frac{\partial \mathbf{A}}{\partial t} = \eta \nabla^2 A\, \mathbf{e_y}. \tag{5.2.7}$$

In 2D the field lines behave as though frozen in to an apparent velocity field $\mathbf{u}$, except at null points where the velocity becomes singular. To derive this velocity

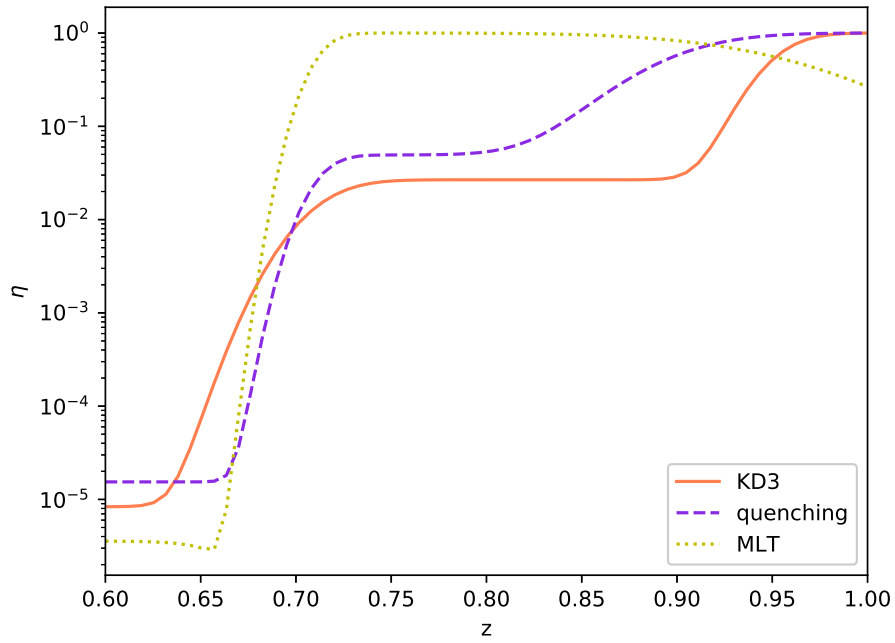

Figure 5.6: Normalized multi-step diffusion profiles used in this thesis, against a log-scale. The solid orange curve is from KD3, the dashed purple curve is the profile which takes into account diffusivity quenching, and the dotted yellow curve is derived from mixing-length theory.

field, write:

$$\frac{\partial \mathbf{A}}{\partial t} = (\mathbf{u} \times \mathbf{B}) + \nabla\psi, \tag{5.2.8}$$

where ψ is a locally single-valued gauge function that can be chosen. Then taking the curl of this expression, we get:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}), \tag{5.2.9}$$

which is the ideal MHD induction equation, valid away from $\mathbf{u} = \mathbf{0}$. Then:

$$(\mathbf{u} \times \mathbf{B}) + \nabla\psi = \eta\nabla^2 A\, \mathbf{e_y}. \tag{5.2.10}$$

Taking the vector product of this equation with $\mathbf{B}$ from the left-hand side we have:

$$\mathbf{u}\,(\mathbf{B}\cdot\mathbf{B}) - \mathbf{B}\,(\mathbf{B}\cdot\mathbf{u}) = \mathbf{B} \times \left(\eta\nabla^2 A\, \mathbf{e_y} - \nabla\psi\right). \tag{5.2.11}$$

Now since Equation 5.2.10 only determines the components of $\mathbf{u}$ perpendicular to

$\mathbf{B}$, the parallel component is arbitrary so we are free to choose $\mathbf{B} \cdot \mathbf{u} = 0$. Hence:

$$\mathbf{u} = \frac{\mathbf{B} \times (\eta \nabla^2 A \, \mathbf{e_y} - \nabla \psi)}{B^2}. \tag{5.2.12}$$

This is well-defined away from null points, i.e. where $B^2 = 0$. We can now choose ψ such that $\dfrac{\partial \psi}{\partial y} = 0$, so

$$u_x = -\frac{B_z \, \eta \nabla^2 A}{B^2}, \tag{5.2.13}$$

and

$$u_z = \frac{B_x \, \eta \nabla^2 A}{B^2}. \tag{5.2.14}$$

We focus on the central axis of $x = 0$, where $B_z \, (0, z) = 0$. Therefore:

$$u_x \, (0, z) = 0, \tag{5.2.15}$$

and

$$\begin{aligned} u_z \, (0, z) &= \frac{\eta \nabla^2 A}{B_x} \\ &= \frac{1}{B_x} \frac{\partial A}{\partial t}, \end{aligned} \tag{5.2.16}$$

since $B^2 = B_x^2$ at $x = 0$. Thus we can track the radial field line velocity at the centre of the domain using Equation 5.2.16. This is useful, because it gives an indication of how flux is transported throughout the convection zone.

For diffusivity in KD3, Yeates & Muñoz-Jaramillo (2013) used Equation 5.2.6 with values of $\eta_c = 10^8 \, \mathrm{cm^2 \, s^{-1}}$, $\eta_0 = 1.6 \times 10^{11} \, \mathrm{cm^2 \, s^{-1}}$, $\eta_s = 6 \times 10^{12} \, \mathrm{cm^2 \, s^{-1}}$, $R_1 = 0.71$, $\Delta_1 = 0.03$, $R_2 = 0.95$ and $\Delta_2 = 0.025$. This profile is displayed as the solid orange curve in Figure 5.6. These parameter choices were originally made so that a full cycle could successfully be simulated. Four snapshots of the simulation using the KD3 profile are shown in Figure 5.7. It is clear from the middle column that there is significantly more flux at the surface than would be expected without radial derivatives, as we saw for the KD3 model in Section 5.1. This is because the relatively low diffusion below $z = 0.9$ does not allow for much diffusive transport, and field lines remain attached to the toroidal field at the base of the convection

zone. This interpretation is supported by the right-hand column, which shows the field line velocity in the z-direction. The field diffuses downwards very slowly due to the gradient in the diffusion profile. Because the field lines in the lower convection zone are essentially fixed in place, movement at the surface is heavily restricted and cancellation at the boundary is limited, resulting in an excess of surface flux. This transpires even though diffusion is stronger near the surface, as indicated by the outward bulging of field lines.

Miesch & Dikpati (2014) presented a self-sustaining 3D kinematic dynamo model (hereafter STABLE), which is described in more detail in Miesch & Teweldebirhan (2016). The main difference between STABLE and KD3 is the flux emergence process: whilst in KD3 a velocity perturbation is applied to the toroidal field and the buoyant rise of flux tubes through the convection zone is explicitly modelled, in STABLE active regions are deposited at the surface, situated above locations of strong toroidal field at the base of the convection zone. We believe that the emergence process in KD3 is more realistic, but multiple examples of self-sustaining dynamo action using STABLE have been presented (Karak & Miesch, 2017, 2018), and Hazra *et al.* (2017) suggested that the method is in principle more suitable for modelling the surface diffusion phase once the BMR has fully emerged.

To demonstrate the different evolution for a disconnected active region such as considered by STABLE, we alter the initial condition slightly from Equation 5.2.4:

$$A_0 = \frac{1}{2}\exp\left(-\frac{x^2 + (z-1)^2}{0.008}\right). \tag{5.2.17}$$

This forms a potential field below the surface, disconnected completely from the base of the convection zone (see top left panel of Figure 5.8).

Figure 5.8 shows four snapshots from the simulation with the original KD3 diffusion profile and disconnected initial condition. Because field lines are no longer connected to the toroidal field at the base of the domain, the weak diffusion no longer plays a role in anchoring field lines in place. This allows for more diffusive transport and cancellation of magnetic flux at the surface. Instead of flux cancellation occurring at

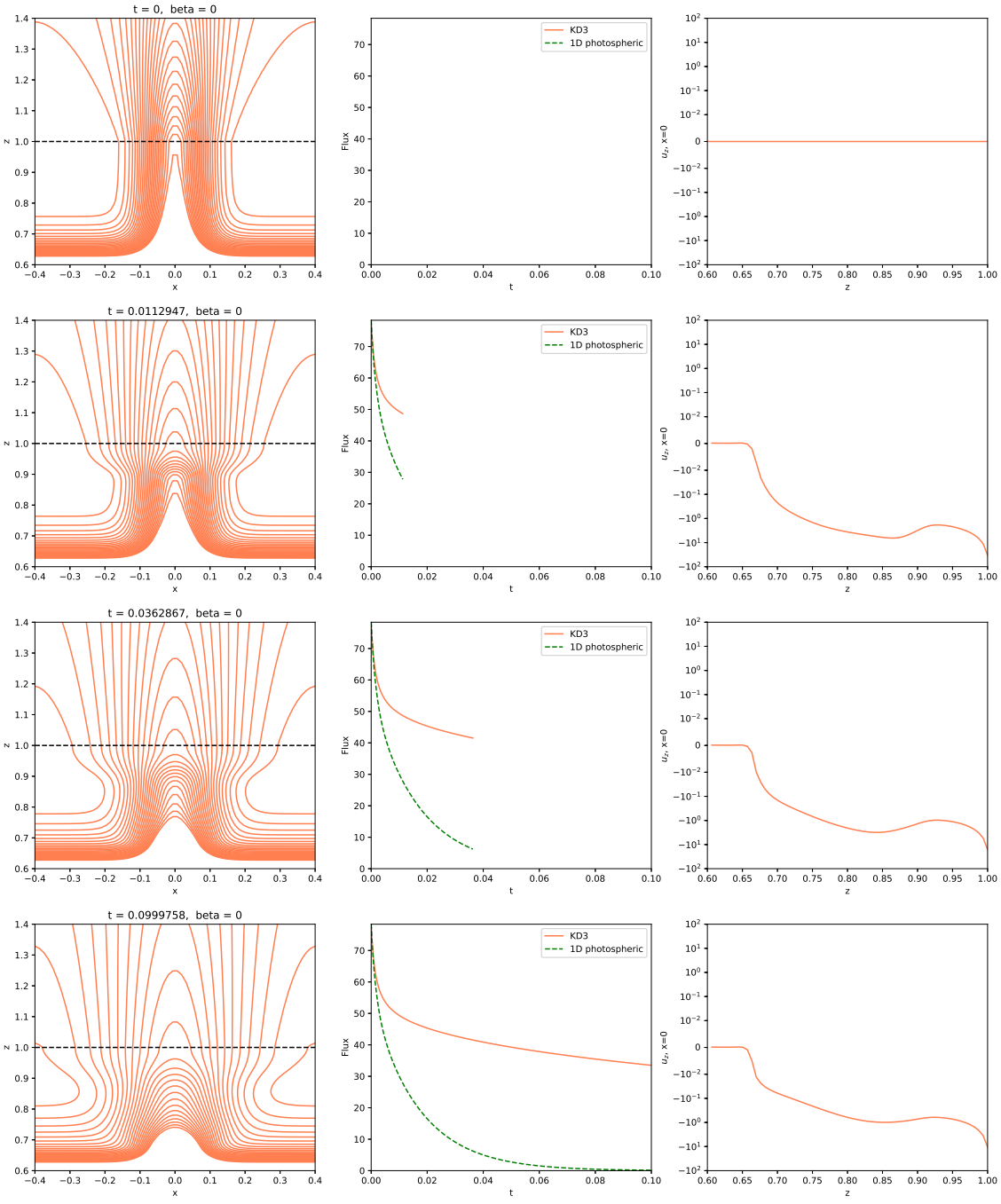

Figure 5.7: Four snapshots of magnetic field lines from the simulation with the diffusion profile used in the original KD3 model. The black dashed line is the top of the domain, above which is shown a potential field extrapolation. The middle column shows the amount of magnetic flux at the surface, compared to a 1D surface model (green dashed line). The right column shows the radial field line velocity at the centre of the domain.

the side boundary after field lines are pushed outwards as before, cancellation takes place between the two polarities of the active region, with a null point appearing at approximately a quarter of the depth of the convection zone, as evidenced by the discontinuities in the right-hand column. The consequence is that there is less surface flux in the early stages of evolution in comparison to the 1D model. The cancellation rate eventually decreases, but we observe in the bottom right panel that there is less surface flux present at the end of the simulation than in the case where the connected initial condition was used. The upshot is that the disconnected region qualitatively provides a better match to the surface than the connected region.

In the presence of strong magnetic fields, turbulent diffusivity can be suppressed (Roberts & Soward, 1975). This 'quenching' can be included in models via a non-linear relationship whereby the diffusion parameter η is scaled by the reciprocal of the square of the magnetic field (e.g. Tobias, 1996; Gilman & Rempel, 2005; Muñoz-Jaramillo *et al.*, 2008; Guerrero *et al.*, 2009). By instead taking the geometric spatiotemporal average over many effective diffusivity profiles, Muñoz-Jaramillo *et al.* (2011) approximated the effect of the dynamically quenched diffusion using a fixed profile in the form of Equation 5.2.6 by using the following parameters: $\eta_c = 10^8\,\mathrm{cm^2\,s^{-1}}$, $\eta_0 = 1.6 \times 10^{11}\,\mathrm{cm^2\,s^{-1}}$, $\eta_s = 3.25 \times 10^{12}\,\mathrm{cm^2\,s^{-1}}$, $R_1 = 0.71$, $\Delta_1 = 0.017$, $R_2 = 0.895$ and $\Delta_2 = 0.051$. This is shown as the dashed purple curve in Figure 5.6, and will henceforth be referred to as the 'quenching profile' for simplicity.

The four snapshots from the simulation using the quenching profile and the original connected initial condition are displayed in Figure 5.9. The field lines diffuse downward initially, but approximately halfway through the simulation the direction of motion changes and the magnetic field starts to diffuse upwards. We note a reduction in the surface flux, presumably because the stronger diffusivity levels extend deeper into the domain and the field lines have more freedom to move, allowing for more diffusive transport. However, we find again that flux cancellation is hindered by the weak diffusion in the lower convection zone, which keeps the field lines attached to

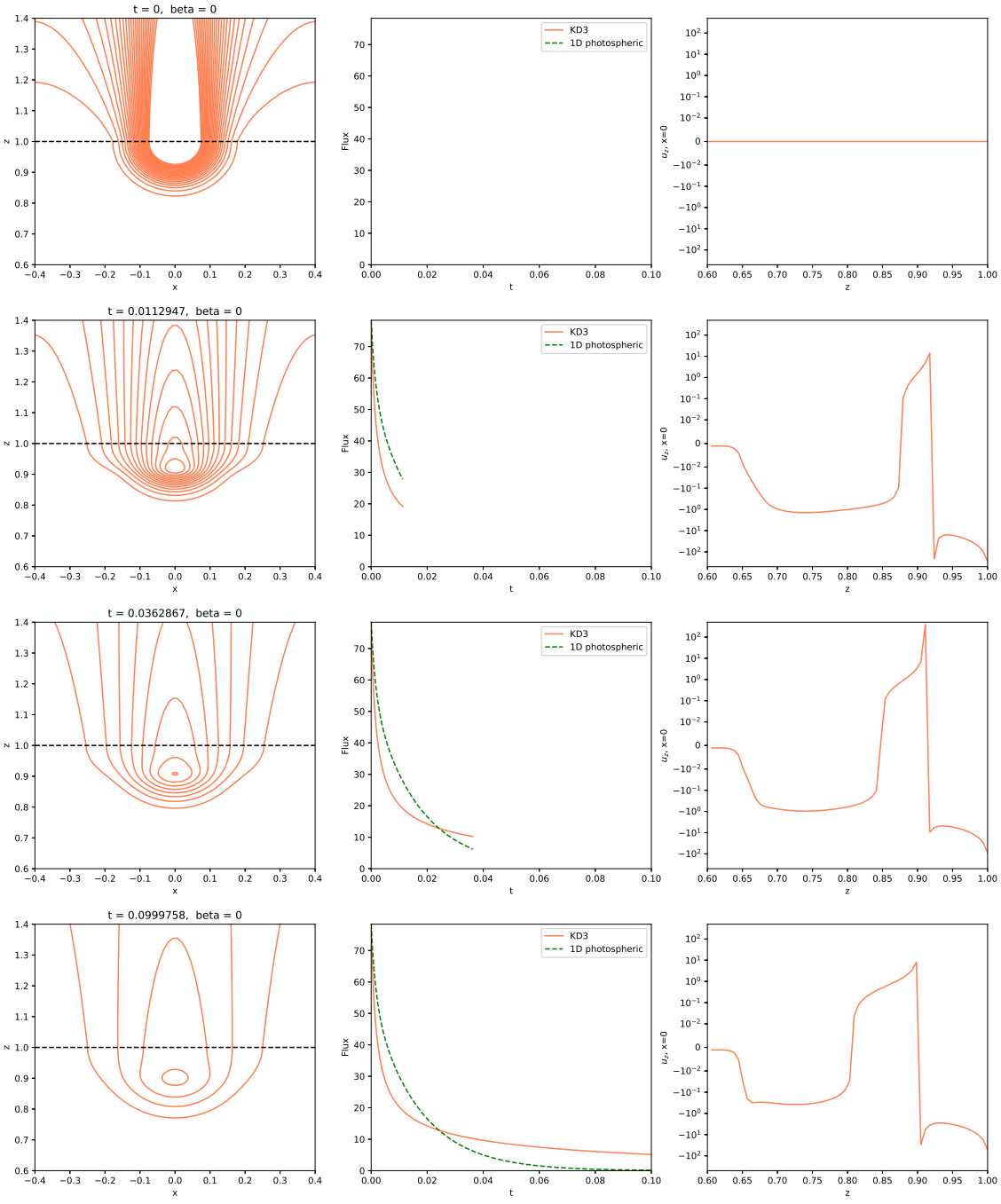

Figure 5.8: Four snapshots of magnetic field lines from the simulation with the KD3 diffusion profile and a disconnected initial magnetic field. The black dashed line is the top of the domain, above which is shown a potential field extrapolation. The middle column shows the amount of magnetic flux at the surface, compared to a 1D surface model (green dashed line). The right column shows the radial field line velocity at the centre of the domain.

the toroidal field.

Figure 5.10 shows four snapshots from the simulation with the quenching diffusion profile and disconnected initial condition. We find that, as for the KD3 profile, flux cancels inwardly because field lines are not connected to the base of the convection zone. However, it diffuses at a much faster rate than the regime with the KD3 diffusion profile (and hence the 1D case), and by the end of the simulation the majority of the surface flux has been cancelled.

The third profile we experiment with is derived from mixing-length theory (MLT; Prandtl, 1925). The mixing-length is the average vertical distance travelled by a fluid element before mixing with the surrounding fluid due to some eddy viscosity, and is usually parametrized as $l \propto \alpha_p H_p$. Here α_p is the mixing-length parameter and H_p is the pressure scale height, $H_p = \dfrac{P}{\rho g}$, where P is pressure, ρ is mass density, and g is gravitational acceleration. These parameters can then be used to estimate the diffusivity. Although the mixing-length model is only a rough approximation, it has been applied to various physical models and has been found to be in agreement with turbulent convection simulations (Chan & Sofia, 1987; Abbett *et al.*, 1997). Muñoz-Jaramillo *et al.* (2011) used the solar interior model of Christensen-Dalsgaard *et al.* (1996) to estimate α_p and H_p and hence the diffusivity profile based on GONG data. The value of diffusion found for the convection zone is up to two orders of magnitude larger than those used in KD3 and other kinematic dynamo simulations in literature. This is because simulated dynamo action has not yet been achieved with such strong diffusion. Muñoz-Jaramillo *et al.* (2011) attempted to reconcile the MLT estimates with numerical values by incorporating diffusivity quenching, leading to the quenching profile above. Nevertheless, a fit to the MLT profile was also made in the form of Equation 5.2.6, with the following resulting parameters: $\eta_c = 10^8\,\mathrm{cm^2\,s^{-1}}$, $\eta_0 = 1.4 \times 10^{13}\,\mathrm{cm^2\,s^{-1}}$, $\eta_s = 10^{10}\,\mathrm{cm^2\,s^{-1}}$, $R_1 = 0.71$, $\Delta_1 = 0.015$, $R_2 = 0.96$ and $\Delta_2 = 0.09$. This profile is the dotted yellow curve shown in Figure 5.6.

Four snapshots from the corresponding simulation are shown in Figure 5.11. With

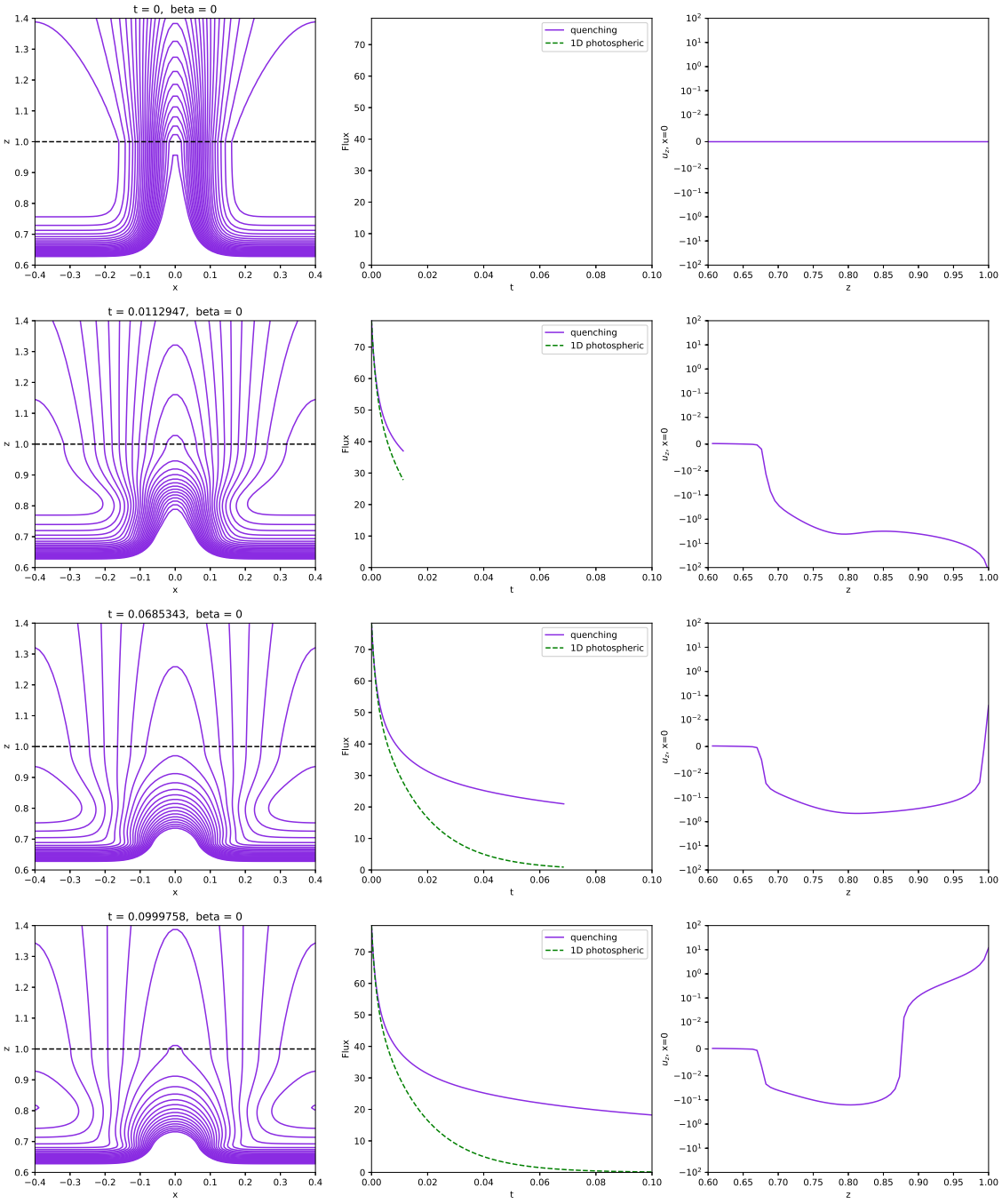

Figure 5.9: Four snapshots of magnetic field lines from the simulation with the diffusion profile that represents diffusivity quenching. The black dashed line is the top of the domain, above which is shown a potential field extrapolation. The middle column shows the amount of magnetic flux at the surface, compared to a 1D surface model (green dashed line). The right column shows the radial field line velocity at the centre of the domain.

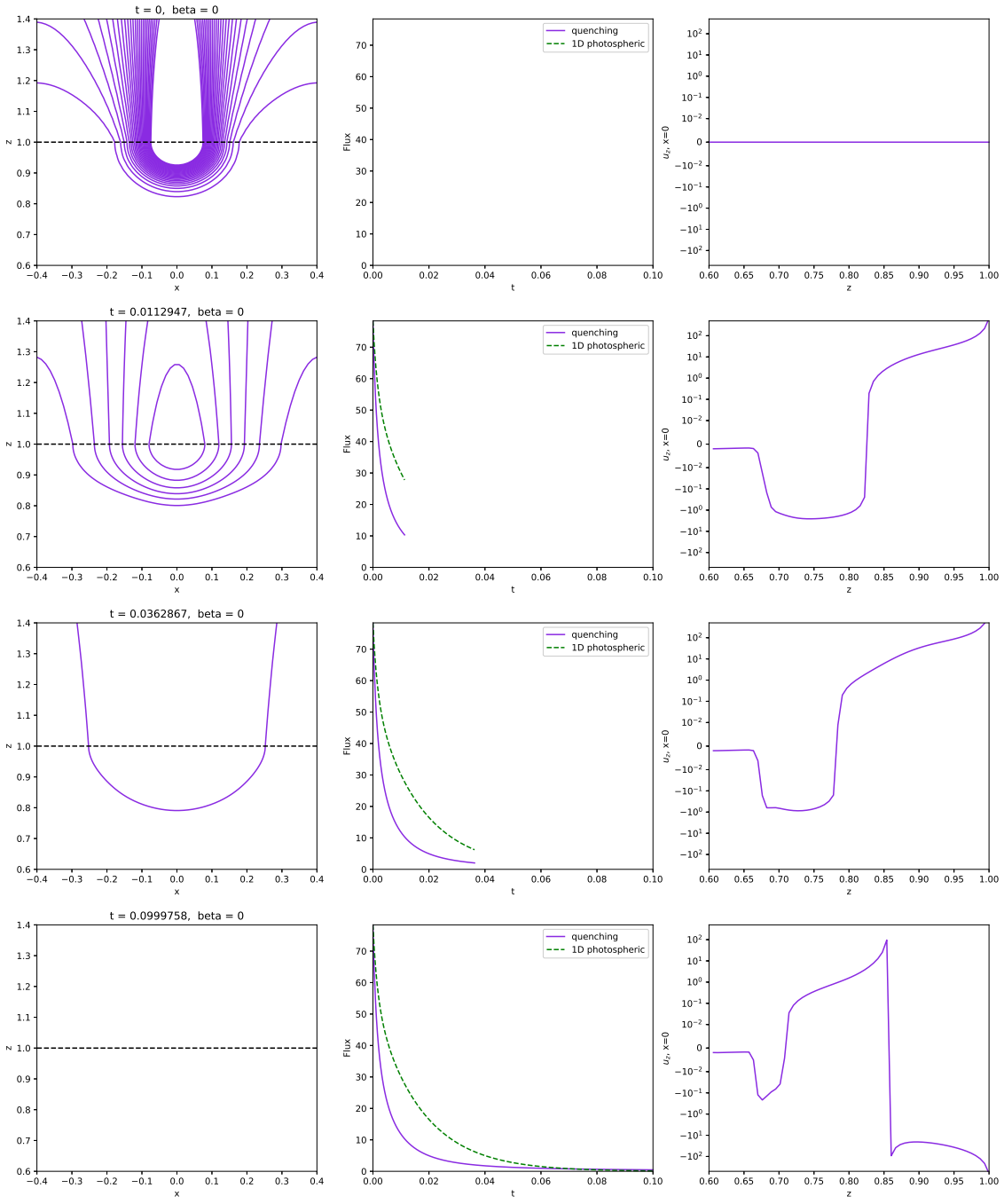

Figure 5.10: Four snapshots of magnetic field lines from the simulation with the quenching diffusion profile and a disconnected initial magnetic field. The black dashed line is the top of the domain, above which is shown a potential field extrapolation. The middle column shows the amount of magnetic flux at the surface, compared to a 1D surface model (green dashed line). The right column shows the radial field line velocity at the centre of the domain.

this diffusion profile, the field initially diffuses downwards before being pushed back up due to the diffusion gradient at the surface. This surface flux then diffuses to the boundary where it cancels. Low diffusivity at the base means the field remains toroidal but a much larger diffusivity throughout the convection zone helps transport flux upwards from as deep as $z = 0.7$. In the right-hand panel of the second row, we see that there is a discontinuity in the velocity, indicating a null point. This is visible in the left-hand panel of the same row, where field lines are being pushed together at the top of the domain due to the reduced diffusivity near the surface and a balance between outward and inward diffusion. At a higher cadence, we observe that this causes the field lines to reconnect. The position of the null initially moves downwards, before changing direction and reaching the surface after approximately a third of the simulation time. After this point, magnetic field diffuses outwards rapidly. In terms of surface flux, this regime is closer to the 1D case than any other two-step profile we test with the connected active region.

Finally, Figure 5.12 shows four snapshots from the simulation with the MLT profile and STABLE initial condition. Because of the strong diffusivity in the bulk of the domain, the field spreads out in the convection zone and diffuses radially outwards due to the reduced diffusivity at the surface. This leads to a surface evolution that matches the 1D case very closely.

It may seem that a good match to the 1D model can be achieved by increasing the diffusion in the 2D model until the surface-flux-versus-time curves lie on top of each other, particularly in the MLT simulation where the general shapes are similar. Tests show that although this does improve the situation somewhat, the match is by no means perfect, especially compared to the performance of the simulation with the MLT profile and disconnected active region.

We now assess the effect of the upper boundary condition on the surface evolution. For this test, we prescribe a constant diffusivity of $\eta = 1$ independent of depth. Four snapshots of the simulation are shown in Figure 5.13. The left-hand column shows magnetic field lines where the upper boundary condition is potential, and the middle

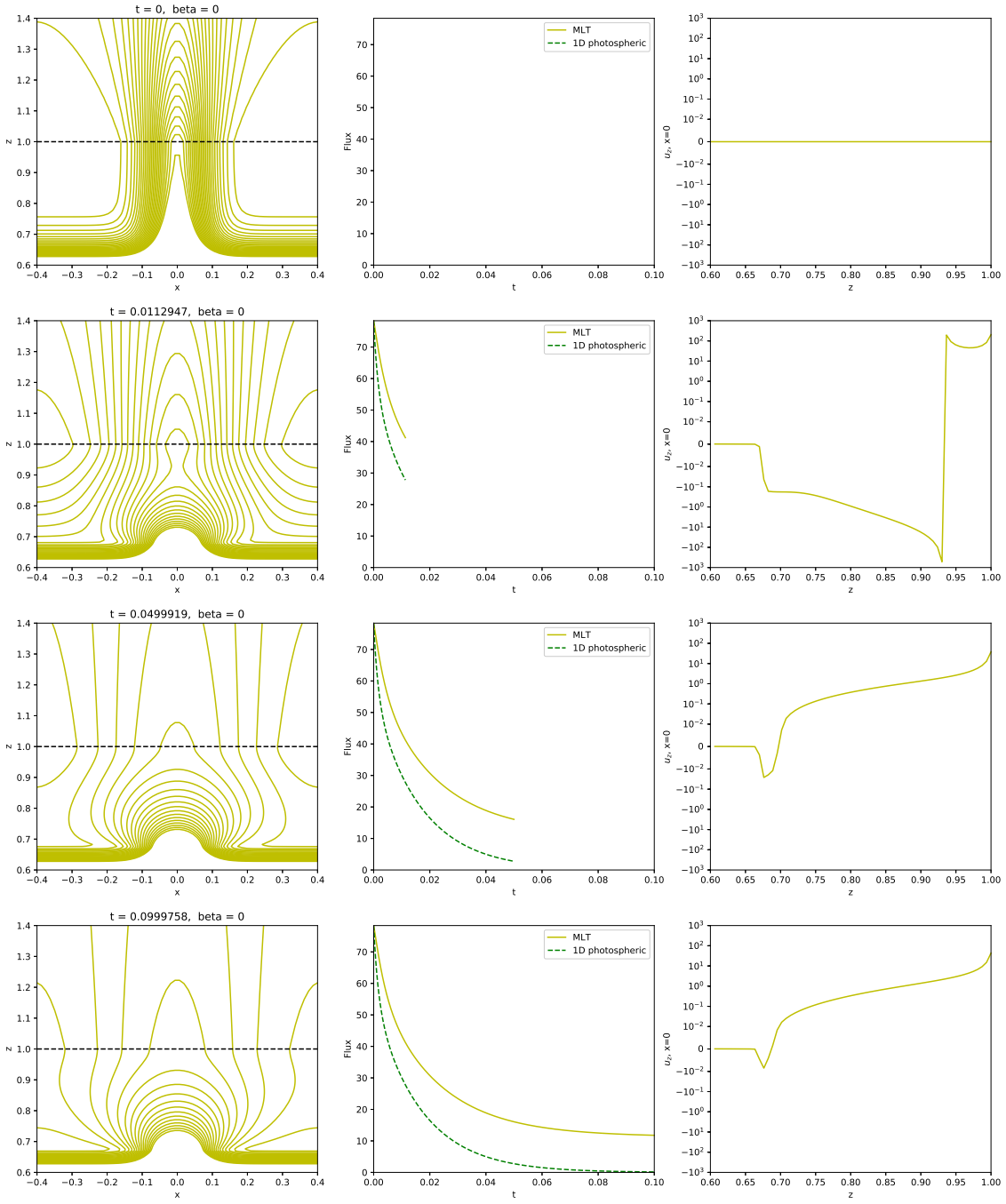

Figure 5.11: Four snapshots of magnetic field lines from the simulation with the diffusion profile derived from MLT. The black dashed line is the top of the domain, above which is shown a potential field extrapolation. The middle column shows the amount of magnetic flux at the surface, compared to a 1D surface model (green dashed line). The right column shows the radial field line velocity at the centre of the domain.

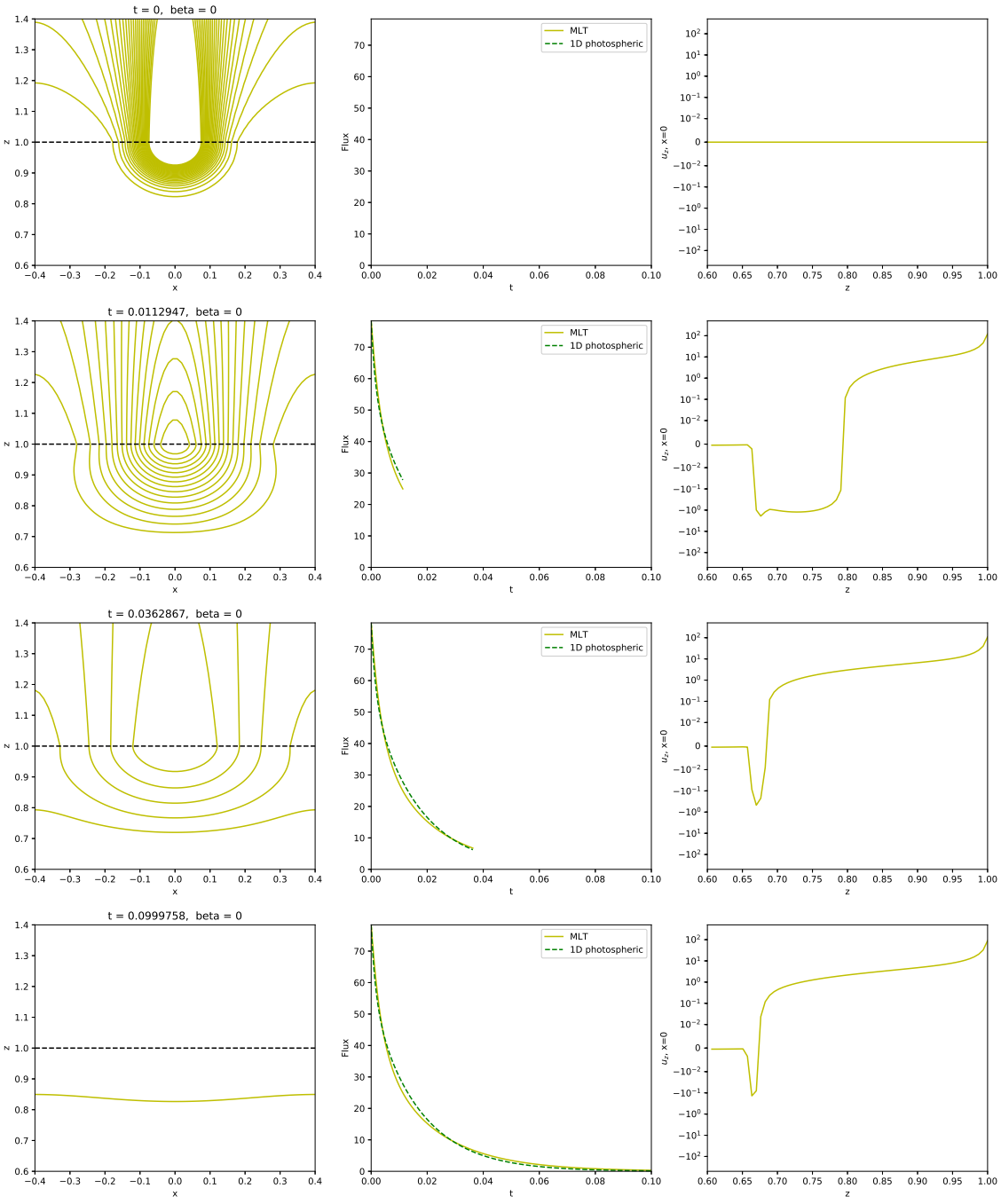

Figure 5.12: Four snapshots of magnetic field lines from the simulation with the MLT diffusion profile and a disconnected initial magnetic field. The black dashed line is the top of the domain, above which is shown a potential field extrapolation. The middle column shows the amount of magnetic flux at the surface, compared to a 1D surface model (green dashed line). The right column shows the radial field line velocity at the centre of the domain.

column shows the field lines where the boundary condition is radial. Qualitative differences are small, but we see in the right-hand column that there is a little too much magnetic flux at the surface in the radial case, compared to the 1D surface model. Conversely, the potential case matches the 1D evolution closely. If $\beta = 1$, we introduce $-\dfrac{\partial B_x}{\partial z}$ into Equation 5.2.3 at the surface which is not present in the radial case. Hence the difference between the two regimes is only situated in the upper quarter of the domain. The enforcement of a radial field at the surface boundary also means that the field lines interact with the periodic boundary later, because they are strictly vertical as opposed to the potential case where cancellation can occur more readily. Since the diffusivity is high throughout the domain, the majority of the flux is diffused out of the convection zone by the time we reach $t = 0.1$ (bottom row of Figure 5.13).

Although the choice of radial or potential-field boundary condition can slightly change the amount of magnetic flux at the surface, the differences are only small, and starker differences arise when we prescribe a more realistic multi-step diffusion profile in place of the constant diffusivity. Further tests show that the small improvement attained by changing boundary condition is the same regardless of the choice of diffusion profile. Interestingly, the 2D model in the constant case provides a good match to the 1D model and explains in part why the MLT profile performs best out of the multi-step profiles we tested: the strong diffusivity allows the magnetic field to diffuse outwards in both cases, the only difference being that the field lines remain attached to the toroidal field in the MLT case due to a weak base diffusion.

The periodic boundary conditions in x can be interpreted as the presence of neighbouring active regions. To check the influence of this inter-region spacing, we tried increasing the width of the domain. This results in more flux present at the surface because it takes longer to diffuse to the boundary and cancel. However, the results above hold qualitatively, and in any case we cannot choose the locations of active region emergence when simulating the evolution of observed BMRs, so varying the width of the domain does not give us significantly deeper insight.

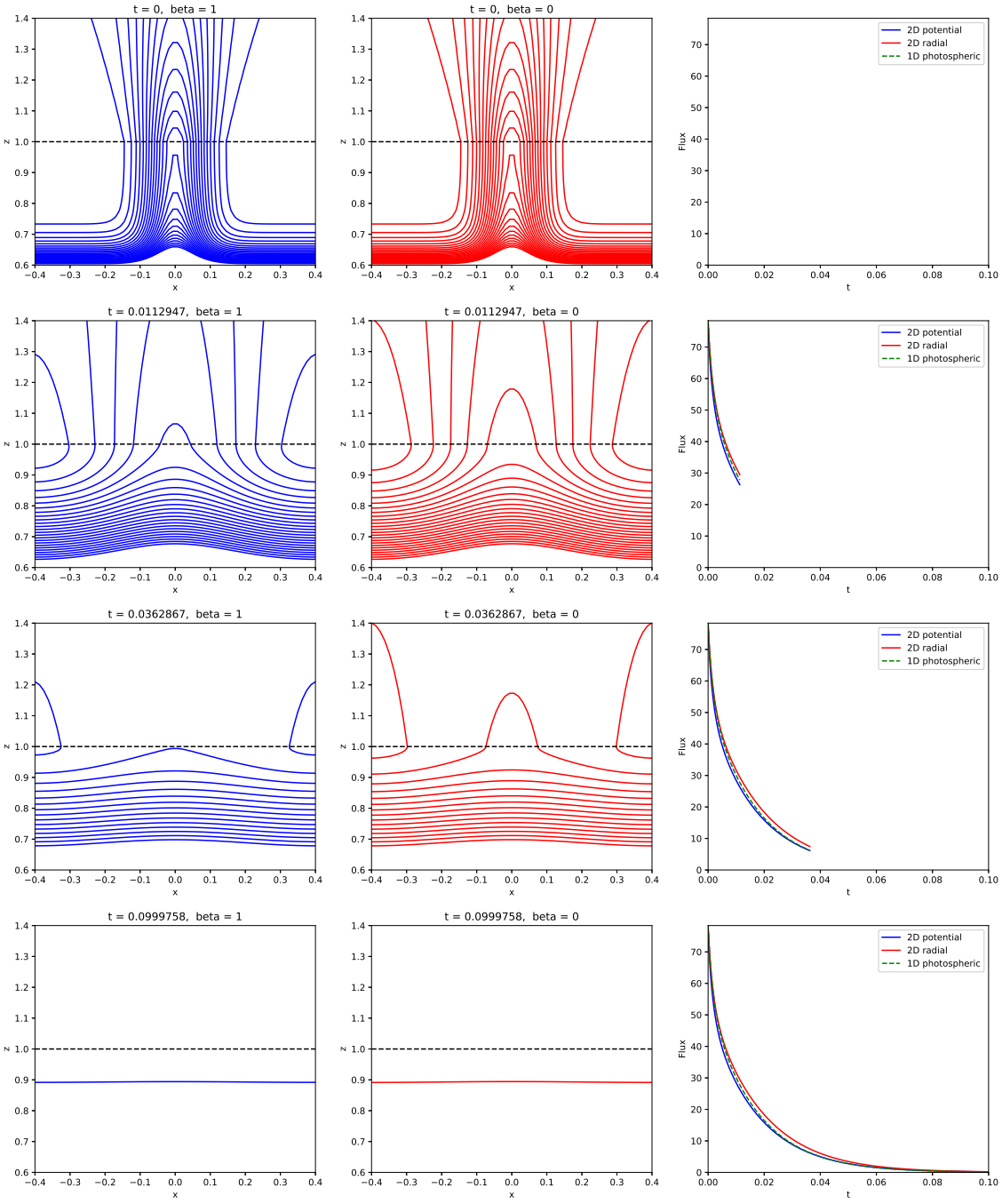

Figure 5.13: Four snapshots of magnetic field lines from the simulation with the potential boundary condition (left column) and radial boundary condition (middle column). The black dashed line is the top of the domain, above which is shown a potential field extrapolation. The right column shows the amount of magnetic flux at the surface in each case, compared to a 1D surface model (green dashed line).

5.3 Effect of diffusivity for a 3D decaying active region

We return to the 3D dynamo model KD3 to test whether the results found in Section 5.2 hold qualitatively here as well. We emerge a single region at 10° latitude with flux 1×10^{22} Mx and a tilt angle of 30°. Because the position of the flux tube during the rising phase depends on differential rotation (Yeates & Muñoz-Jaramillo, 2013), we turn on rotation until the region has emerged after 25 days, at which time the velocity perturbation is turned off. A snapshot of the system is taken on that day, and all subsequent experiments are run from time of emergence, regardless of whether differential rotation is included.

Differential rotation takes the form of Charbonneau *et al.* (1999), which was derived from observations:

$$\begin{aligned}\Omega\left(r,\theta\right) =& \Omega_c + \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{r - R_0}{\Delta_0}\right)\right] \\ & \cdot \left[\Omega_E - \Omega_c + \left(\Omega_p - \Omega_E\right)\left(C\cos^2\theta + \left(1 - C\right)\cos^4\theta\right)\right], \end{aligned} \tag{5.3.1}$$

where $\Omega_c = 2.714\,34 \times 10^{-6}\,\mathrm{s}^{-1}$, $\Omega_E = 2.9531 \times 10^{-6}\,\mathrm{s}^{-1}$, $\Omega_p = 2.073\,45 \times 10^{-6}\,\mathrm{s}^{-1}$, $C = 0.483$, $R_0 = 0.7\,R_\odot$ and $\Delta_0 = 0.025\,R_\odot$.

For meridional flow we first define the following stream function:

$$\begin{aligned}\Psi\left(r,\theta\right) =& \frac{-v_0\left(r - R_p\right)}{7.633\,r\sin\theta}\sin\left(\pi\frac{r - R_p}{R_\odot - R_p}\right)\exp\left(-\left(\frac{r - R_1}{\Gamma}\right)^2\right) \\ & \cdot \left(1 - \exp\left(-1.5\theta^2\right)\right)\left(1 - \exp\left[1.8\left(\theta - \frac{\pi}{2}\right)\right]\right), \end{aligned} \tag{5.3.2}$$

where $R_p = 0.62\,R_\odot$, $R_1 = 0.1125\,R_\odot$, $\Gamma = 3.47 \times 10^8\,\mathrm{m}$ and $v_0 = 20\,\mathrm{m\,s}^{-1}$. Note that this can be written as a function of r multiplied by a function of θ. Then the meridional circulation is given by

$$\mathbf{u_m} = \frac{1}{\rho\left(r\right)}\nabla \times \left(\Psi\left(r,\theta\right)\mathbf{e}_\phi\right), \tag{5.3.3}$$

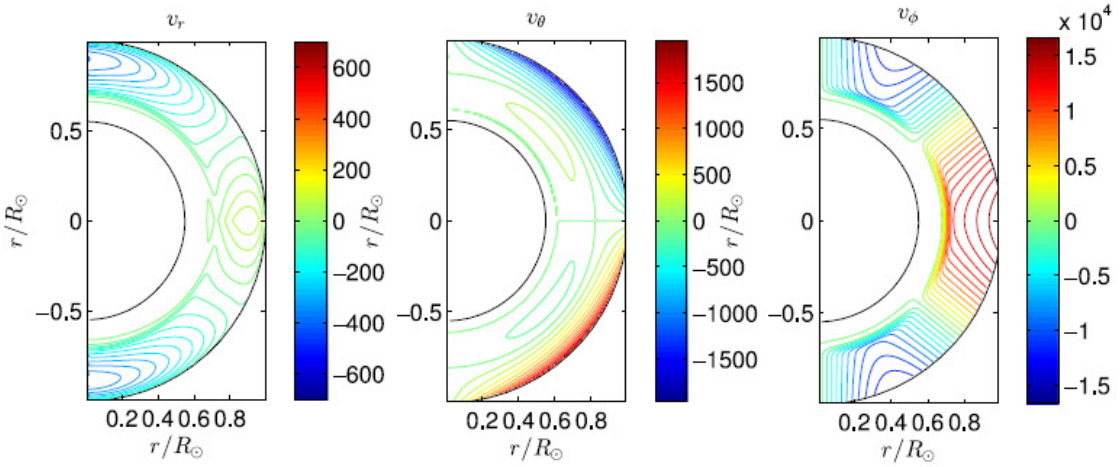

Figure 5.14: Depictions of velocity components used in this chapter, taken from Yeates & Muñoz-Jaramillo (2013). The left and middle panels primarily show meridional circulation, and the right panel shows differential rotation. Colour-scale units are in $\mathrm{cm\,s^{-1}}$.

where $\rho\,(r) = \left(\dfrac{R_\odot}{r} - 0.95\right)^{\frac{3}{2}}$ is the radial density profile. This produces a single cell circulation profile with a poleward flow at the surface, as observed, and an equatorward branch that penetrates the tachocline (Nandy & Choudhuri, 2002). Whilst it is likely that the meridional flow consists of a more complex form than just a single cell, Hazra *et al.* (2014) showed that solar behaviour is reproduced as long as an equatorward return flow is present. The velocity components are shown in Figure 5.14. Diffusivity is now given by $\eta_s\,\eta(r/R_\odot)$ using Equation 5.2.6, where $\eta_s = 6\times10^{12}\,\mathrm{cm^2\,s^{-1}}$.

We use a grid resolution of $\Delta_\phi = \dfrac{2\pi}{384}$ and $\Delta_r = \dfrac{0.45\,R_\odot}{48}$, and initial and boundary conditions are the same as used by Yeates & Muñoz-Jaramillo (2013): the bottom boundary condition at $r = 0.55\,R_\odot$ is a perfectly conducting core, i.e. $\dfrac{\partial(rB_\theta)}{\partial r} = \dfrac{\partial(rB_\phi)}{\partial r} = 0$. The upper boundary condition is radial, although we expect from Section 5.2 that changing to a potential-field boundary condition would have a negligible effect on the flux evolution. The initial condition is created by emerging a single BMR from a belt of toroidal field at the base of the convection zone of the form:

$$\mathbf{B} = \frac{B_0}{2}\left(\mathrm{erf}\left(\frac{r-R_7}{\Delta_8}\right) - \mathrm{erf}\left(\frac{r-R_8}{\Delta_8}\right)\right)\mathbf{e}_\phi, \tag{5.3.4}$$

with $R_7 = 0.66\,R_\odot$, $R_8 = 0.74\,R_\odot$, $\Delta_8 = 0.018\,R_\odot$ and $B_0 = 2.5 \times 10^3\,\mathrm{G}$. At the surface, we do not prescribe any initial magnetic field, aside from that of the emerged BMR.

We run the model for 9000 days from the time of emergence with no post-emergence differential rotation, using a different diffusion profile from Section 5.2 each time. This best reflects the scenario modelled in the simplified 2D diffusion model in Section 5.2. The top-left panel of Figure 5.15 shows the unsigned surface flux from each of these simulations, while the bottom-left panel shows the polar fluxes. As we found in the 2D model (Section 5.2), the KD3 profile (orange) restricts cancellation, resulting in a vast excess of flux at the surface. Poleward transport is also slower in this case; the northern polar flux still hasn't reached its peak after the full 9000 days.

Qualitatively, the other profiles also exhibit the same behaviour as in the 2D model. The MLT profile (yellow) provides a more rapid decay of flux, and the quenching profile (purple) lies somewhere between the other two. Unfortunately, even after 9000 simulation days, there is $\sim 1 \times 10^{22}\,\mathrm{Mx}$ of unsigned magnetic flux at the surface in all three regimes – much more than in the SFT model as shown by the dotted blue (without exponential decay) and black (with a decay term of $\tau = 10$ years) curves. Whilst the decay term in the SFT model makes only a very small difference in the total unsigned surface flux, its impact at the poles is more evident, acting as a sink for the polar flux which is not otherwise possible in the SFT model. Although the peak strength of the northern polar field is weaker in the MLT case than the SFT model, it occurs at a similar time and the shape of the profile is close to that of the SFT model when exponential decay is included.

If we do not turn off differential rotation after 25 days, we obtain the curves plotted in the middle column of Figure 5.15. On comparing the first two columns of Figure 5.15, it is clear that the shearing caused by differential rotation aids the cancellation of flux at the surface. By the end of the simulation, the amount of flux at the surface eventually decreases to a similar level to that of the SFT model in all regimes,

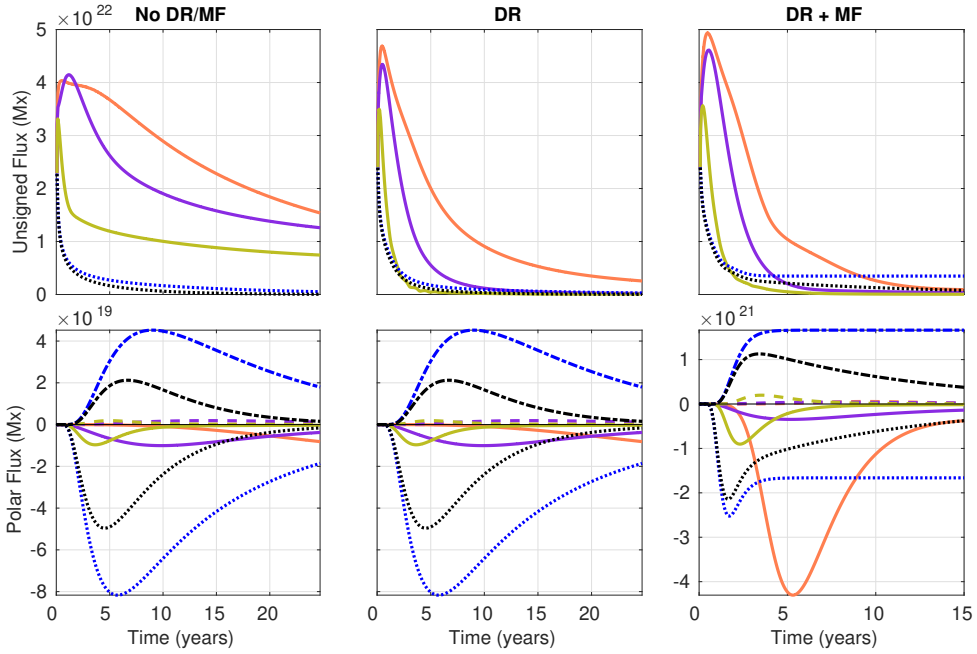

Figure 5.15: Top row: Unsigned surface flux from 3D simulations of a single active region under the effects of diffusion only (left), diffusion plus differential rotation (middle), and diffusion, rotation and meridional flow (right), using the KD3 diffusion profile (orange), quenching profile (purple) and MLT profile (yellow). The equivalent 2D SFT flux is shown in the dotted blue (without decay) and black (with decay) curves. Bottom row: Northern polar flux (solid and dotted lines) and southern polar flux (dashed and dash-dotted lines) from the same simulations.

although the evolution of unsigned flux is still different in each case. Even with the aid of differential rotation, none of the diffusion profiles are able to produce results that match the SFT model, and the qualitative proximity of each curve to the SFT curve found in previous tests is preserved. Note that the signed polar flux remains the same regardless of whether rotation is included. This is because differential rotation drops out when we average over longitude.

If we also include meridional circulation, we obtain the results plotted in the right-hand column of Figure 5.15. This is the same setup as described in Section 5.1 but starting from a different time. We show only the first fifteen years of simulation time to focus on the main evolution; after this point, the curves remain at the same level asymptotically. Qualitatively, the three profiles give similar results. As with the rotation-only case, by the end of the simulation there is a similar amount of surface flux in the KD3 simulations as in the SFT simulations. The meridional flow transports flux to the poles on a shorter timescale than the other two cases, resulting in a much stronger polar field. The inability of the KD3 diffusion profile to remove magnetic flux from the surface means that the peak polar flux is stronger than that of the SFT simulation. We observe that once again the MLT profile produces the closest qualitative match to the surface-only model.

5.4 Effect of diffusivity on a 3D full-cycle simulation

Yeates & Muñoz-Jaramillo (2013) demonstrated a simulation of a full solar cycle using BMR data from Solar Cycle 23. However, this was not systematically calibrated to observations. It can be seen in Figure 5.16 (or equivalently Figure 12 of Yeates & Muñoz-Jaramillo, 2013) that the magnetic field is too strong and poleward surges are too slow compared to the optimal butterfly diagram found in Section 2.2, shown in the lower panel of Figure 5.16, which was calibrated against observations. The

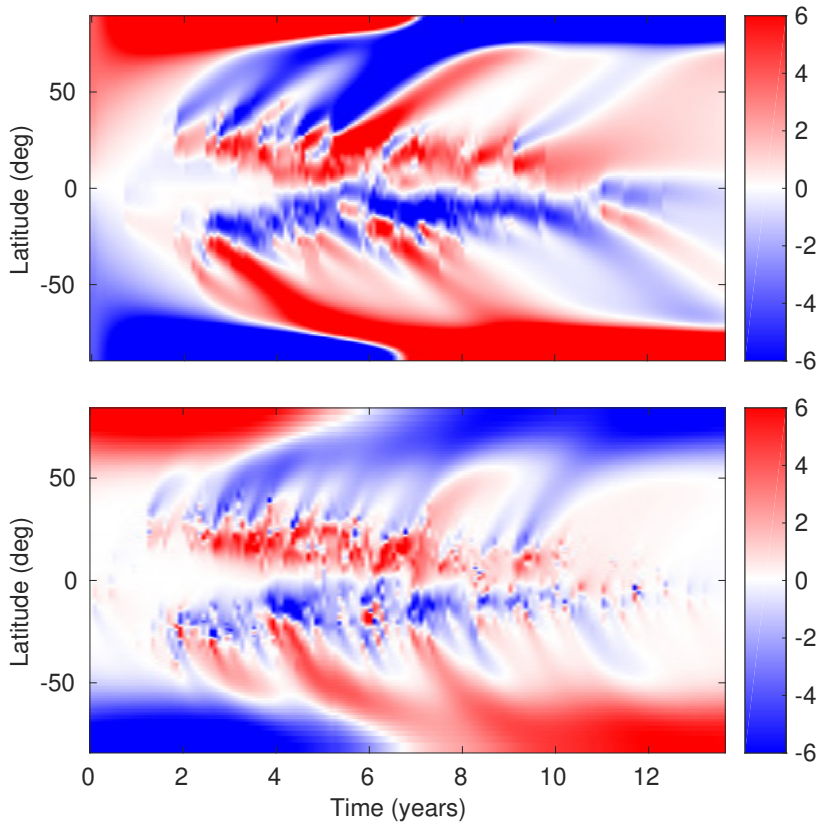

Figure 5.16: Top: Simulation of Cycle 23 from Yeates & Muñoz-Jaramillo (2013). Bottom: Optimal butterfly diagram of Cycle 23 from Section 2.2 of this thesis.

active regions across the full solar cycle behave similarly to the individual region in Figure 5.2. We repeat this 3D simulation of Cycle 23 but replace the original diffusion profile (orange curve in Figure 5.6) with the quenching and mixing-length theory profiles (purple and yellow curves in Figure 5.6 respectively). Ideally we would like to run a full solar cycle using disconnected active regions, but calculating the interior potential field is time-consuming and we leave this for future work.

Equation 5.3.4 again defines the initial toroidal field, but now we try $B_0 = 250\,\mathrm{G}$. An initial dipolar field is given by

$$\mathbf{B} = \nabla \times (A_\phi \mathbf{e}_\phi)\,, \tag{5.4.1}$$

where

$$A_\phi = B_d \frac{\sin\theta}{r^3} \left(\frac{r - 0.7\,R_\odot}{0.3\,R_\odot} \right), \tag{5.4.2}$$

and $A_\phi = 0$ for $r < 0.7\,R_\odot$ (Jouve *et al.*, 2008). The field strength is set as

$B_d = -0.008\, B_0$.

We run the simulation for 5000 days, using observed BMRs of Cycle 23 from NSO/Kitt Peak as input data (Yeates *et al.*, 2007). The unsigned surface flux and signed polar flux for the simulation of Yeates & Muñoz-Jaramillo (2013) are shown by the orange curves (top and bottom respectively) in Figure 5.17. The blue and black curves show the 2D SFT simulation without and with decay respectively for comparison. Note that here the regions are automatically extracted from NSO magnetograms and are not the BMRs used in KD3, otherwise all surface parameters are the same and we use $B_{\rm par} = 39.8\,\mathrm{G}$ (see Chapter 2). Therefore the cause of the discrepancy is no longer obvious. In particular, Yeates & Muñoz-Jaramillo (2013) state that the magnetic flux in KD3 can be scaled arbitrarily by simply changing the strength of the toroidal field, so the difference in scale here is not significant.

The purple profiles in Figure 5.17 correspond to the simulation where the quenching diffusivity profile has been used. If all parameters other than the diffusivity profile are fixed, it is evident that not enough magnetic flux reaches the surface, and the polar field is barely able to reverse. To combat this, we increase the strength of the initial toroidal field by an order of magnitude. This provides a stronger source from which active regions can develop, thereby increasing the amount of flux at the photosphere. This is demonstrated by the purple curve in the top panel of Figure 5.18. Here, the total surface flux peaks earlier than the original simulation. In the bottom panel, we see that the polar field reverses at a similar time to the original case, albeit with a reduced strength throughout the simulation. Nevertheless, the toroidal field appears to be strong enough to produce more regions as a subsequent cycle (top panel of Figure 5.19) if we were to continue the simulation. The bottom panel of Figure 5.19 shows the surface butterfly diagram of the same simulation. While it is suboptimal, it displays observable features of the solar cycle and a more realistic distribution and transport of magnetic flux than before. A future task is to calibrate other parameters in the model against observations while keeping the quenching profile fixed.

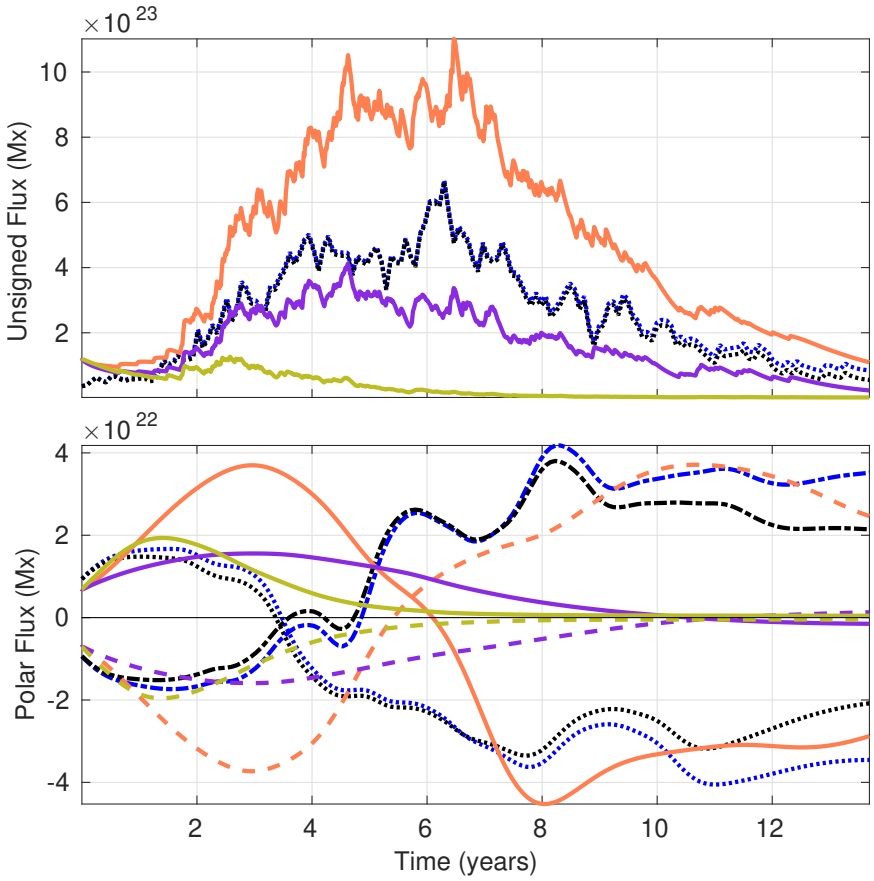

Figure 5.17: Top: Unsigned surface flux from 3D simulations of Cycle 23 using the KD3 diffusion profile (orange), quenching profile (purple) and MLT profile (yellow). The equivalent 2D SFT flux is shown in the dotted blue (without decay) and black (with decay) curves. Bottom: Northern polar flux (solid and dotted lines) and southern polar flux (dashed and dash-dotted lines) from the same simulations.

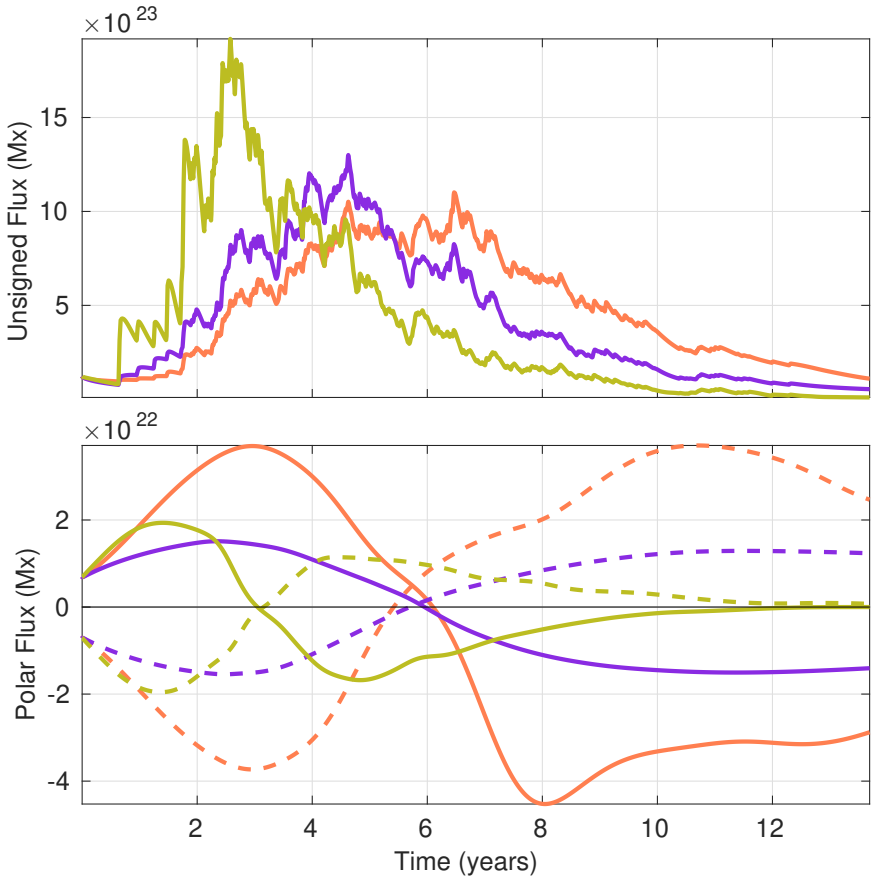

Figure 5.18: Top: Unsigned surface flux from 3D simulations of Cycle 23 using the KD3 diffusion profile (orange), quenching profile (purple) and MLT profile (yellow), but where the initial toroidal field has been strengthened by one and two orders of magnitude for the latter two respectively. Bottom: Northern polar flux (solid line) and southern polar flux (dashed line) from the same simulations.

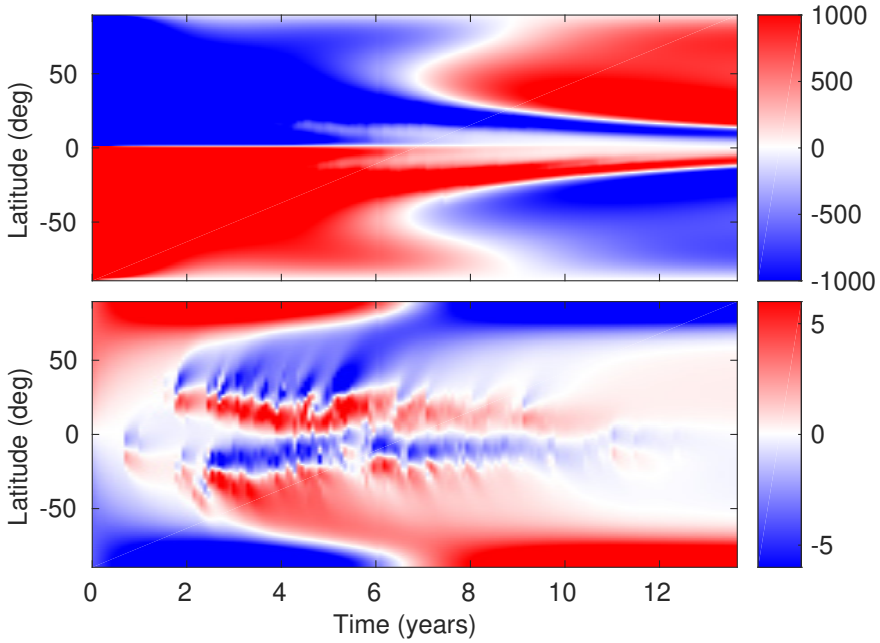

Figure 5.19: Top: Toroidal field at the base of the convection zone from a 3D simulation of Cycle 23 using the quenching profile and a strengthened initial toroidal field. Bottom: Radial magnetic field at the surface from the same simulation.

Ideally we would like to be able to simulate Cycle 23 using the diffusion profile derived from mixing-length theory, because this gave the closest match to the surface-only evolution in Sections 5.2 and 5.3. Figure 5.17 shows that even less flux emerges at the surface in this case, because the diffusion in the convection zone is too strong and kills off the majority of rising flux tubes. Even when the initial toroidal field is increased by an order of magnitude, it rapidly diffuses and so no regions are able to emerge after a few years.

When the toroidal field is increased by another order of magnitude, the flux still decays too rapidly, as shown by the yellow curve in Figure 5.18. However, we now observe polar field reversal, although very early in the cycle, and the bottom panel of Figure 5.20 shows that the surface evolution during the first few years of the cycle appears to be sun-like. The top panel of Figure 5.20 shows that no new toroidal field is created. This occurs in all simulations when the MLT diffusivity profile is used and is one reason why dynamos have thus far been unable to accommodate the diffusion profile derived from mixing-length theory.

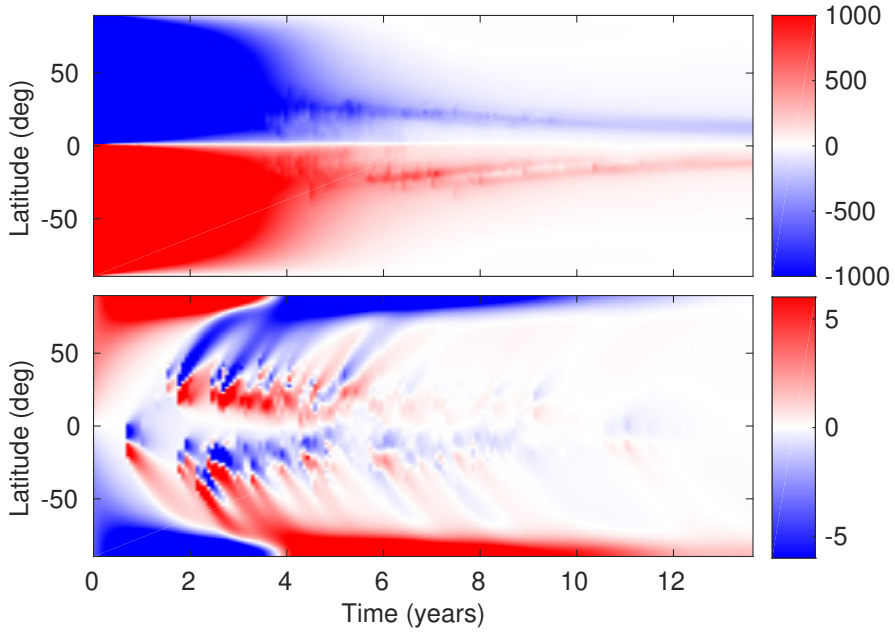

Figure 5.20: Top: Toroidal field at the base of the convection zone from a 3D simulation of Cycle 23 using the MLT profile and a strengthened initial toroidal field. Bottom: Radial magnetic field at the surface from the same simulation.

Scaling the MLT profile by a factor of 0.5 allows significantly more flux to emerge at the surface, but it is still not enough on its own to sustain the dynamo. However, if we also shift the location of the low-diffusivity step in the MLT profile up so that the toroidal field is stored in a region of low diffusion (i.e. set $R_1 = 0.74$ and $\Delta_1 = 0.024$), we find that the field survives for longer and more flux can reach the surface. However, although more new toroidal field starts to appear at the base of the convection zone for the next cycle, it is still too weak, and the polar field at the surface still reverses too early. In summary, increasing the diffusivity to the level required for a realistic surface evolution is not on its own sustainable in a full-cycle simulation, because the high diffusivity removes too much flux from the system.

5.5 Conclusions

In this chapter our aim was to investigate and fix a discrepancy in surface evolution between a 3D dynamo model and 2D surface flux transport model. Suspecting that

the problem was probably due to radial diffusion, we used a simple 2D model of the convection zone to quickly assess the effects of diffusion profile, initial condition and boundary conditions. We simultaneously solved a 1D diffusion model to create a reduced-dimension analogue of the original problem.

We found that weak diffusivity at the base of the domain meant that field lines were unable to detach from the toroidal field and were held fixed in place. This meant that movement of magnetic flux at the surface was severely limited, restricting diffusion and cancellation. Because of its relatively reduced diffusivity in the convection zone, using the original KD3 diffusion profile in particular left a significant excess of unsigned flux at the surface. However, this can be reduced by increasing the diffusivity in the convection zone so that diffusive transport is more effective, as demonstrated by diffusion profiles derived from diffusivity quenching estimates and mixing-length theory. There was an improvement in surface flux when potential rather than radial boundary conditions were used, but the difference was negligible compared to improvements gained by changing the diffusion profile.

We also tested an initial condition where the magnetic field was disconnected from the toroidal field at the base of the convection zone. Field lines were no longer fixed in place so diffusion was very effective. Using the MLT profile with this initial condition produced a very good match between the surface magnetic flux and the flux from the 1D model. Evolving a disconnected active region provided a better fit to the 1D surface evolution in general, although in the case of the quenching profile, diffusion acted too rapidly.

Based on these results, we believe that the source of the discrepancy between the SFT model and the surface component of KD3 arises from a combination of prescribing a weak diffusivity in the bulk of the convection zone and maintaining a connected subsurface structure of active regions, and that these become manifest through the radial diffusion term.

We then repeated some of the experiments in KD3, and found qualitatively similar results: the MLT diffusion profile produced the closest match in surface flux evolution

to the SFT model, whilst the quenching and KD3 profiles performed second and third best respectively. The MLT profile also induced a small improvement in polar flux evolution.

Finally, we made an attempt of simulating Cycle 23 with KD3 using the quenching and MLT profiles. After increasing the strength of the initial toroidal field, we were able to present an acceptable representation of Cycle 23 using the quenching profile. However, using the preferred MLT profile was largely unsuccessful, and the toroidal field decayed away without any new toroidal field being produced to act as the seed for the next cycle. While this profile worked best for simulating a single region, applying the strong diffusivity throughout the convection zone meant that the cycle was not able to sustain itself for a sufficiently long time, and very little toroidal flux was being produced for the next cycle. The implication of this is that in order to *accurately* recreate the solar cycle using a 3D diffusion-dominated dynamo model, it may be that emerged regions should at some point be disconnected from the underlying toroidal field (Schüssler & Rempel, 2005), in order for the evolution of surface flux to match that of an SFT model.

Chapter 6

Conclusions and future work

In this thesis we investigated methods for optimizing parameters of solar magnetic cycle models. In Chapter 2 we developed a 1D surface flux transport model and applied a genetic algorithm to this model to search for optimal parameters. We also applied the algorithm to a 2D surface flux transport model which uses an automated method to extract specific active region shape from magnetograms. In Chapter 3 we used Bayesian emulation to calibrate parameters for the 2D model, and in Chapter 4 we used the 2D model along with its optimal parameters to simulate active regions individually and obtain properties such as axial dipole moment, latitude and flux. Finally, in Chapter 5 we assessed the effect of radial diffusion at the surface of a 3D dynamo model, compared to the 2D surface flux transport model. We now summarise the conclusions from each chapter:

Chapter 2: Parameter optimization for surface flux transport models.

- A 1D surface flux transport model is developed which uses observed bipolar magnetic regions as source terms.
- Optimal parameters and 'acceptable ranges' for parameters are successfully found using a genetic algorithm. Parameter relationships are also inferred.

- Observational uncertainty is incorporated to add statistical significance to the results.
- The genetic algorithm is also applied to a 2D surface flux transport model which extracts specific active region shape from magnetograms.
- Optimal parameters for both models are in good agreement with observations.
- The 2D model accommodates the observed meridional flow better than the 1D model, and does not require an additional exponential decay term. Its automated assimilation technique allows for quick simulations of different datasets.
- There exists variation between optimal parameters from different solar cycles.
- Global search techniques are not common in this research area and are often overlooked. Work in this chapter builds on similar studies by Lemerle *et al.* (2015) and Lemerle & Charbonneau (2017), but instead the algorithm is applied to two distinct models and different solar cycles.

Chapter 3: The Bayesian approach.

- Bayesian emulation is applied to the same optimization problem as Chapter 2, specifically for the 2D model. Such a technique has never been used before in the area of Solar Physics.
- Emulation provides a more efficient and informative parameter search. The multi-dimensional parameter space can be regularly visualized, and regions of space ruled out based on a quantity called 'implausibility'.
- Results are in excellent agreement with Chapter 2, and significantly fewer model runs are required.
- Efficiency of emulators increases with model complexity. Therefore a more sophisticated emulator would be an appropriate tool for optimizing the 3D kinematic dynamo model in Chapter 5.

Chapter 4: How many active regions are necessary to predict the solar dipole moment?

- Active regions and their properties from Solar Cycles 21 to 24 are extracted from observed magnetograms. Each region is simulated individually so its contribution to the dipole moment can be calculated.
- Large contributors can have a drastic effect on the end-of-cycle dipole moment and hence subsequent cycle amplitude.
- The cumulative effect of smaller contributors is also important.
- Emergence latitude primarily determines the dipole moment contribution of a region, but there are secondary dependencies on initial magnetic flux and dipole moment.
- In particular, all significant contributions are from regions above a certain size.
- This improves on work by Jiang *et al.* (2015), who used an empirical relation involving active region emergence properties to determine dipole moment contributions, rather than a simulation.

Chapter 5: Three-dimensional kinematic dynamo model.

- There is a discrepancy between the evolution of a 2D surface flux transport model and the surface component of a 3D dynamo model.
- This is caused by radial diffusion which is absent in the surface-only model.
- Tests using a simple 2D diffusion model show that a strong diffusivity is required in the convection zone in order for the surface flux to match.
- Alternatively, disconnecting the active region from the toroidal field at the base of convection zone provides a more accurate surface flux evolution.
- The effect of changing surface boundary condition appears to be negligible.

- Repeating the experiments in the 3D dynamo model confirms the results.
- It does not yet appear to be possible, in this particular model, to maintain a full realistic solar cycle with a strong convection zone diffusion and active regions initially connected to the underlying toroidal field.
- The aforementioned discrepancy seems to be largely unexplored – at least in the context of flux transport models – so the work in this chapter is original and hopefully will be impactful.

6.1 Potential future work

The adaptability of the 2D model provides a wide scope of possible future directions for optimization. One such direction is testing variability between different measuring instruments to ascertain whether inconsistent literature results could simply be due to the choice of observatory or satellite. This comes with the issue of either deciding on or computing an appropriate value for the assimilation threshold $B_{\rm par}$ for different datasets. Another future possibility that takes advantage of the model's assimilation technique is to optimize multiple cycles at the same time. We have shown that there exists variation in parameters between cycles, so a single optimal parameter set for more than one cycle might be less realistic. An alternative method would be to treat each cycle separately, coupled only at each cycle minimum, where the final profile of the previous cycle becomes the initial profile of the next. However, using this method we found in Chapter 4 that errors in early cycles can propagate through the simulations and affect later cycles, and that using a single parameter set actually improves the overall match between observed and simulated axial dipole moment.

Our methodology assumes a static meridional flow. The inclusion of a time-varying meridional flow in the optimization could significantly alter results, however parametrizing time-dependence without introducing too many parameters is not a trivial procedure. On the other hand, large-scale inflows towards active regions were first

observed by Gizon *et al.* (2001), and Cameron & Schüssler (2012) proposed that these flows were at least partially responsible for variation of meridional flow over the solar cycle. Indeed, Martin-Belda & Cameron (2016) found that the inflows increased the effect of flux cancellation and also reduced the latitudinal separation of polarities, thereby decreasing the axial dipole moment contribution of a bipolar region. This process weakens the polar field in the same way that a time-dependent meridional flow can, and although we have not accounted for inflows in this study, it is an option under consideration for future work. A possible technique for simulating the effect of active region inflows without having to calculate a flow pattern explicitly is using a flux-dependent diffusion parameter whereby the presence of a strong magnetic field quenches diffusion (e.g. Muñoz-Jaramillo *et al.*, 2011).

We would also like to accurately simulate solar cycles earlier than Cycle 21. Unfortunately, magnetograms are not available for these cycles. However, an alternative possibility is to combine chromospheric plage data from Ca II K line spectroheliograms with polarity information from historic sunspot magnetic field measurements to create pseudo-magnetograms (Pevtsov *et al.*, 2016). These could subsequently be used as assimilation data for the 2D SFT model in order to gain deeper insight into the behaviour of earlier cycles, particularly the polar field (Virtanen *et al.*, 2017, 2018). It may also be possible to optimize parameters for these historical cycles against other datasets such as filament observations, but this is ongoing work.

Remarkably, even a relatively basic Bayesian emulator was successful. One would hope that a more sophisticated emulator would be even more accurate and efficient. For example, the functions $\beta_i g_i(x)$ in Equation 3.1.1 were chosen to be constant as opposed to more complicated functions like low-order polynomials, which might more accurately describe the global behaviour of the model. Furthermore, we considered all variables to be 'active', i.e. all parameters influence the output $f(x)$. However we may decide that some variables are less important than others in determining the output, and only construct the emulator based on those which have a greater influence. This enhances efficiency by reducing dimensions, and would therefore

be more useful when calibrating more complex models than the 2D SFT model. Active variables can be chosen using statistical model-fitting criteria such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC). Further modifications and improvements are discussed in Vernon *et al.* (2010).

In Chapter 5 we only varied the diffusion profile and initial condition in the 3D dynamo model. Perhaps a way to find a working dynamo with connected active regions would be to vary the other parameters (e.g. meridional flow profile, turbulent pumping). As discussed throughout this thesis, the end goal is to systematically calibrate the full set of parameters in KD3 against the real solar cycle, so we would like to use the work in Chapters 2, 3 and 5 to add some constraints to the optimization problem. It may be that modelling the solar cycle is still possible using active regions that remain connected to the toroidal field at the base of the convection zone.

A subsequent future target would be to achieve a self-sustaining dynamo using this model, where the locations of emerging regions are automatically chosen based on the strength of the toroidal field and not to match the observed regions necessarily. In particular, we would like achieve this whilst ensuring the surface magnetic field behaviour is sun-like. This remains an open research problem for 3D B-L dynamo models and work is being done on this topic by other groups.

For example, the 3D B-L model of Miesch & Dikpati (2014) has been shown to successfully produce self-sustaining dynamo action in multiple tests of increasing complexity, such as adding axisymmetric and convective flows (Miesch & Teweldebirhan, 2016; Hazra & Miesch, 2018), operating in a regime where diffusion dominates over advection (Hazra *et al.*, 2017), and experimenting with solar cycle variability (Karak & Miesch, 2017, 2018). However, this model uses an idealized spot-deposition algorithm and does not explicitly model the emergence process, which, aside from probably being less realistic, comes with a small number of problems. For instance, the interior structure of the magnetic field must be defined (in this case a potential field is assumed), and the depletion of flux from the toroidal field needs to be taken into account in an appropriate way, which is by no means a trivial (or, indeed,

resolved) task. Furthermore, using such an algorithm allows for the alteration of active region parameters, such as emergence time and delay, size, magnetic flux, latitude and tilt angle. In many of the studies listed above, the flux of each region was amplified above observed solar values in order to ensure supercritical dynamo solutions.

Kumar *et al.* (2018) and Kumar *et al.* (2019) presented another 3D B-L dynamo model, which uses a similar emergence process to that of Yeates & Muñoz-Jaramillo (2013). Whilst promising self-sustaining dynamo simulations were performed, the model displayed some shortcomings, such as an overlap between consecutive cycles. In like fashion to the model of Miesch & Dikpati (2014), the magnetic flux of active regions was too strong, and a detailed comparison to the behaviour of the solar magnetic cycle was not performed, leaving lots of avenues for future work.

Bibliography

Abbett, W. P., Beaver, M., Davids, B., Georgobiani, D., Rathbun, P. & Stein, R. F. 1997 Solar Convection: Comparison of Numerical Simulations and Mixing-Length Theory. *ApJ* **480**, 395–399.

Andrianakis, I., Vernon, I. R., McCreesh, N., McKinley, T. J., Oakley, J. E., Nsubuga, R. N., Goldstein, M. & White, R. G. 2015 Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda. *PLOS Computational Biology* **11** (1), 1–18.

Babcock, H. W. 1961 The Topology of the Sun's Magnetic Field and the 22-YEAR Cycle. *ApJ* **133**, 572.

Babcock, H. W. & Babcock, H. D. 1955 The Sun's Magnetic Field, 1952-1954. *ApJ* **121**, 349.

Baumann, I. 2005 *Magnetic Flux Transport on the Sun*. Copernicus GmbH.

Baumann, I., Schmitt, D. & Schüssler, M. 2006 A necessary extension of the surface flux transport model. *A&A* **446**, 307–314.

Baumann, I., Schmitt, D., Schüssler, M. & Solanki, S. K. 2004 Evolution of the large-scale magnetic field on the solar surface: A parameter study. *A&A* **426**, 1075–1091.

Schulte in den Bäumen, H., Moran, D., Lenzen, M., Cairns, I. & Steenge, A. 2014 How severe space weather can disrupt global supply chains. *Natural Hazards and Earth System Sciences* **14** (10), 2749–2759.

Beer, J. 2000 Long-term indirect indices of solar variability. *Space Sci. Rev.* **94**, 53–66.

Beer, J., Tobias, S. & Weiss, N. 1998 An Active Sun Throughout the Maunder Minimum. *Sol. Phys.* **181**, 237–249.

Benz, A. O. 2017 Flare Observations. *LRSP* **14**, 2.

Bower, R. G., Vernon, I., Goldstein, M., Benson, A. J., Lacey, C. G., Baugh, C. M., Cole, S. & Frenk, C. S. 2010 The parameter space of galaxy formation. *MNRAS* **407**, 2017–2045.

Braun, D. C. & Fan, Y. 1998 Helioseismic Measurements of the Subsurface Meridional Flow. *ApJL* **508**, L105–L108.

Brown, G. 1986 Working Group "A" Report: Long-Term Solar Activity Predictions. In *Solar-Terrestrial Predictions* (ed. P. A. Simon, G. Heckman & M. A. Shea), p. 1.

Bushby, P. & Mason, J. 2004 Solar dynamo: Understanding the solar dynamo. *A&G* **45** (4), 7.

Byrd, R. H., Hribar, M. E. & Nocedal, J. 1999 An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization* **9** (4), 877–900.

Cameron, R. H., Dasi-Espuig, M., Jiang, J., Işık, E., Schmitt, D. & Schüssler, M. 2013 Limits to solar cycle predictability: Cross-equatorial flux plumes. *A&A* **557**, A141.

Cameron, R. H., Jiang, J., Schmitt, D. & Schüssler, M. 2010 Surface Flux Transport Modeling for Solar Cycles 15-21: Effects of Cycle-Dependent Tilt Angles of Sunspot Groups. *ApJ* **719**, 264–270.

Cameron, R. H., Jiang, J. & Schüssler, M. 2016 Solar Cycle 25: Another Moderate Cycle? *ApJL* **823**, L22.

Cameron, R. H., Jiang, J., Schüssler, M. & Gizon, L. 2014 Physical causes of solar cycle amplitude variability. *J. Geophys. Res. (Space Phys.)* **119**, 680–688.

Cameron, R. H. & Schüssler, M. 2012 Are the strengths of solar cycles determined by converging flows towards the activity belts? *A&A* **548**, A57.

Cameron, R. H. & Schüssler, M. 2016 The turbulent diffusion of toroidal magnetic flux as inferred from properties of the sunspot butterfly diagram. *A&A* **591**, A46.

Carrington, R. C. 1859 Description of a Singular Appearance seen in the Sun on September 1, 1859. *MNRAS* **20**, 13–15.

Carrington, R. C. 1863 *Observations of the spots on the Sun from November 9, 1853, to March 24, 1861, made at Redhill*. London [etc.] Williams and Norgate.

Chan, K. L. & Sofia, S. 1987 Validity tests of the mixing-length theory of deep convection. *Science* **235**, 465–467.

Charbonneau, P. 2002*a* An introduction to genetic algorithms for numrical optimization. *NCAR Tech. Note, NCAR/TN-450+IA (Boulder, CO: National Center for Atmospheric Research)* **1**.

Charbonneau, P. 2002*b* Release notes for `PIKAIA` 1.2. *NCAR Tech. Note, NCAR/TN-451+STR (Boulder, CO: National Center for Atmospheric Research)* **1**.

Charbonneau, P. 2014 Solar Dynamo Theory. *ARA&A* **52**, 251–290.

Charbonneau, P., Christensen-Dalsgaard, J., Henning, R., Larsen, R. M., Schou, J., Thompson, M. J. & Tomczyk, S. 1999 Helioseismic Constraints on the Structure of the Solar Tachocline. *ApJ* **527**, 445–460.

Charbonneau, P. & Knapp, B. 1995 A user's guide to `PIKAIA` 1.0. *NCAR Tech. Note, NCAR/TN-418+IA (Boulder, CO: National Center for Atmospheric Research)* **1**.

Chatterjee, P., Nandy, D. & Choudhuri, A. R. 2004 Full-sphere simulations of a circulation-dominated solar dynamo: Exploring the parity issue. *A&A* **427**, 1019–1030.

Chen, P. F. 2011 Coronal Mass Ejections: Models and Their Observational Basis. *LRSP* **8**, 1.

Cheung, M. C. M. & Isobe, H. 2014 Flux Emergence (Theory). *LRSP* **11**, 3.

Choudhuri, A. R., Schussler, M. & Dikpati, M. 1995 The solar dynamo with meridional circulation. *A&A* **303**, L29.

Christensen-Dalsgaard, J., Dappen, W., Ajukov, S. V., Anderson, E. R., Antia, H. M., Basu, S., Baturin, V. A., Berthomieu, G., Chaboyer, B., Chitre, S. M., Cox, A. N., Demarque, P., Donatowicz, J., Dziembowski, W. A., Gabriel, M., Gough, D. O., Guenther, D. B., Guzik, J. A., Harvey, J. W., Hill, F., Houdek, G., Iglesias, C. A., Kosovichev, A. G., Leibacher, J. W., Morel, P., Proffitt, C. R., Provost, J., Reiter, J., Rhodes, Jr., E. J., Rogers, F. J., Roxburgh, I. W., Thompson, M. J. & Ulrich, R. K. 1996 The Current State of Solar Modeling. *Science* **272**, 1286–1292.

Courant, R., Friedrichs, K. & Lewy, H. 1928 Über die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen* **100**, 32–74.

Craig, P. S., Goldstein, M., Seheult, A. H. & Smith, J. A. 1997 Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments. In *Case Studies in Bayesian Statistics* (ed. C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi & N. D. Singpurwalla), pp. 37–93. Springer New York, New York, NY.

Currin, C., Mitchell, T., Morris, M. & Ylvisaker, D. 1991 Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments. *Journal of the American Statistical Association* **86** (416), 953–963.

Dasi-Espuig, M., Solanki, S. K., Krivova, N. A., Cameron, R. & Peñuela, T. 2010 Sunspot group tilt angles and the strength of the solar cycle. *A&A* **518**, A7.

De Moortel, I. & Browning, P. 2015 Recent advances in coronal heating. *Philosophical Transactions of the Royal Society of London Series A* **373**, 20140269–20140269.

DeVore, C. R., Boris, J. P. & Sheeley, Jr., N. R. 1984 The concentration of the large-scale solar magnetic field by a meridional surface flow. *Sol. Phys.* **92**, 1–14.

DeVore, C. R., Sheeley, Jr., N. R., Boris, J. P., Young, Jr., T. R. & Harvey, K. L. 1985 Simulations of magnetic-flux transport in solar active regions. *Sol. Phys.* **102**, 41–49.

Dikpati, M., de Toma, G., Gilman, P. A., Arge, C. N. & White, O. R. 2004 Diagnostics of Polar Field Reversal in Solar Cycle 23 Using a Flux Transport Dynamo Model. *ApJ* **601**, 1136–1151.

Dikpati, M., Gilman, P. A. & Ulrich, R. K. 2010 Physical Origin of Differences Among Various Measures of Solar Meridional Circulation. *ApJ* **722**, 774–778.

Dormand, J. R. & Prince, P. J. 1980 A family of embedded Runge-Kutta formulae. *Journal of Computational and Applied Mathematics* **6**, 19–26.

Durney, B. R. 1995 On a Babcock-Leighton dynamo model with a deep-seated generating layer for the toroidal magnetic field. *Sol. Phys.* **160**, 213–235.

Durney, B. R. 1997 On a Babcock-Leighton Solar Dynamo Model with a Deep-seated Generating Layer for the Toroidal Magnetic Field. IV. *ApJ* **486**, 1065–1077.

Durrant, C. J., Kress, J. M. & Wilson, P. R. 2001 The Evolution of Trailing Plumes from Active Regions. *Sol. Phys.* **201**, 57–69.

Durrant, C. J., Turner, J. & Wilson, P. R. 2002 Bipolar Magnetic Fields Emerging at High Latitudes. *Sol. Phys.* **211**, 103–124.

Durrant, C. J. & Wilson, P. R. 2003 Observations and Simulations of the Polar Field Reversals in Cycle 23. *Sol. Phys.* **214**, 23–39.

Eddy, J. A. 1976 The Maunder Minimum. *Science* **192**, 1189–1202.

Emilio, M., Kuhn, J. R., Bush, R. I. & Scholl, I. F. 2012 Measuring the Solar Radius from Space during the 2003 and 2006 Mercury Transits. *ApJ* **750**, 135.

Fan, Y. 2009 Magnetic Fields in the Solar Convection Zone. *LRSP* **6**, 4.

Gilman, P. A. & Rempel, M. 2005 Concentration of Toroidal Magnetic Field in the Solar Tachocline by η-Quenching. *ApJ* **630**, 615–622.

Gizon, L. & Duvall, T. L. 2004 Solar-cycle variations in the spectrum of supergranulation. In *Multi-Wavelength Investigations of Solar Activity* (ed. A. V. Stepanov, E. E. Benevolenskaya & A. G. Kosovichev), *IAU Symposium*, vol. 223, pp. 41–44.

Gizon, L., Duvall, Jr., T. L. & Larsen, R. M. 2001 Probing Surface Flows and Magnetic Activity with Time-Distance Helioseismology. In *Recent Insights into the Physics of the Sun and Heliosphere: Highlights from SOHO and Other Space Missions* (ed. P. Brekke, B. Fleck & J. B. Gurman), *IAU Symposium*, vol. 203, p. 189.

Goldstein, M. 1999 *Bayes linear analysis*. Wiley, New York. In Kotz, S., Read, C.B. & Banks, D.L: Encyclopaedia of Statistical Sciences Update Volume 3, 29-34.

Goldstein, M., Seheult, A. & Vernon, I. 2013 *Assessing Model Adequacy*. John Wiley & Sons, Ltd, Chichester, UK. In Wainwright, J., Mulligan, M.: Environmental Modelling: Finding Simplicity in Complexity, 2nd edn.

Goldstein, M. & Wooff, D. A. 2007 *Bayes Linear Statistics: Theory and Methods*. John Wiley, Chichester.

Gopalswamy, N., Mäkelä, P., Yashiro, S. & Akiyama, S. 2018 Long-term solar activity studies using microwave imaging observations and prediction for cycle 25. *Journal of Atmospheric and Solar-Terrestrial Physics* **176**, 26–33.

Green, J. L., Boardsen, S., Odenwald, S., Humble, J. & Pazamickas, K. A. 2006 Eyewitness reports of the great auroral storm of 1859. *Advances in Space Research* **38**, 145–154.

Guerrero, G. & de Gouveia Dal Pino, E. M. 2008 Turbulent magnetic pumping in a Babcock-Leighton solar dynamo model. *A&A* **485**, 267–273.

Guerrero, G., Dikpati, M. & de Gouveia Dal Pino, E. M. 2009 The Role of Diffusivity Quenching in Flux-transport Dynamo Models. *ApJ* **701**, 725–736.

Guerrero, G., Rheinhardt, M., Brandenburg, A. & Dikpati, M. 2012 Plasma flow versus magnetic feature-tracking speeds in the Sun. *MNRAS* **420**, L1–L5.

Guerrero, G., Smolarkiewicz, P. K., de Gouveia Dal Pino, E. M., Kosovichev, A. G. & Mansour, N. N. 2016 Understanding Solar Torsional Oscillations from Global Dynamo Models. *ApJL* **828**, L3.

Hale, G. E. 1908 On the Probable Existence of a Magnetic Field in Sun-Spots. *ApJ* **28**, 315.

Hale, G. E. 1924 The Law of Sun-Spot Polarity. *Proceedings of the National Academy of Science* **10**, 53–55.

Hart, A. B. 1954 Motions in the Sun at the photospheric level. IV. The equatorial rotation and possible velocity fields in the photosphere. *MNRAS* **114**, 17.

Hathaway, D. H. 2010 The Solar Cycle. *LRSP* **7**, 1.

Hathaway, D. H. & Rightmire, L. 2010 Variations in the Sun's Meridional Flow over a Solar Cycle. *Science* **327**, 1350.

Hathaway, D. H. & Upton, L. 2014 The solar meridional circulation and sunspot cycle variability. *J. Geophys. Res. (Space Phys.)* **119**, 3316–3324.

Hathaway, D. H. & Upton, L. A. 2016 Predicting the amplitude and hemispheric asymmetry of solar cycle 25 with surface flux transport. *J. Geophys. Res. (Space Phys.)* **121**, 10.

Hawkes, G. & Berger, M. A. 2018 Magnetic Helicity as a Predictor of the Solar Cycle. *Sol. Phys.* **293**, 109.

Hazra, G., Choudhuri, A. R. & Miesch, M. S. 2017 A Theoretical Study of the Build-up of the Sun's Polar Magnetic Field by using a 3D Kinematic Dynamo Model. *ApJ* **835**, 39.

Hazra, G., Karak, B. B. & Choudhuri, A. R. 2014 Is a Deep One-cell Meridional Circulation Essential for the Flux Transport Solar Dynamo? *ApJ* **782** (2), 93.

Hazra, G. & Miesch, M. S. 2018 Incorporating Surface Convection into a 3D Babcock-Leighton Solar Dynamo Model. *ApJ* **864**, 110.

Hodgson, R. 1859 On a curious Appearance seen in the Sun. *MNRAS* **20**, 15–16.

Howard, R. 1983 Torsional Oscillations of the Sun. *Sol. Phys.* **82**, 437–437.

Howard, R. F. 1991 Axial tilt angles of sunspot groups. *Sol. Phys.* **136**, 251–262.

Howe, R. 2009 Solar Interior Rotation and its Variation. *LRSP* **6**, 1.

Howe, R., Davies, G. R., Chaplin, W. J., Elsworth, Y., Basu, S., Hale, S. J., Ball, W. H. & Komm, R. W. 2017 The Sun in transition? Persistence of near-surface structural changes through Cycle 24. *MNRAS* **470**, 1935–1942.

Jackiewicz, J., Serebryanskiy, A. & Kholikov, S. 2015 Meridional Flow in the Solar Convection Zone. II. Helioseismic Inversions of GONG Data. *ApJ* **805**, 133.

Jackson, S. E., Vernon, I., Liu, J. & Lindsey, K. 2018 Bayesian uncertainty analysis establishes the link between the parameter space of a complex model of hormonal crosstalk in Arabidopsis root development and experimental measurements. *arXiv e-prints* p. arXiv:1801.01538.

Jiang, J., Cameron, R. H., Schmitt, D. & Schüssler, M. 2011 The solar magnetic field since 1700. II. Physical reconstruction of total, polar and open flux. *A&A* **528**, A83.

Jiang, J., Cameron, R. H., Schmitt, D. & Schüssler, M. 2013 Can Surface Flux Transport Account for the Weak Polar Field in Cycle 23? *Space Sci. Rev.* **176**, 289–298.

Jiang, J., Cameron, R. H. & Schüssler, M. 2014 Effects of the Scatter in Sunspot Group Tilt Angles on the Large-scale Magnetic Field at the Solar Surface. *ApJ* **791**, 5.

Jiang, J., Cameron, R. H. & Schüssler, M. 2015 The Cause of the Weak Solar Cycle 24. *ApJL* **808**, L28.

Jiang, J. & Cao, J. 2018 Predicting solar surface large-scale magnetic field of Cycle 24. *Journal of Atmospheric and Solar-Terrestrial Physics* **176**, 34–41.

Jiang, J., Işık, E., Cameron, R. H., Schmitt, D. & Schüssler, M. 2010 The Effect of Activity-related Meridional Flow Modulation on the Strength of the Solar Polar Magnetic Field. *ApJ* **717**, 597–602.

Jouve, L., Brun, A. S., Arlt, R., Brandenburg, A., Dikpati, M., Bonanno, A., Käpylä, P. J., Moss, D., Rempel, M., Gilman, P., Korpi, M. J. & Kosovichev, A. G. 2008 A solar mean field dynamo benchmark. *A&A* **483**, 949–960.

Karak, B. B. & Miesch, M. 2017 Solar Cycle Variability Induced by Tilt Angle Scatter in a Babcock-Leighton Solar Dynamo Model. *ApJ* **847**, 69.

Karak, B. B. & Miesch, M. 2018 Recovery from Maunder-like Grand Minima in a Babcock-Leighton Solar Dynamo Model. *ApJL* **860**, L26.

Kennedy, M. C. & O'Hagan, A. 2001 Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63** (3), 425–464.

Komm, R., González Hernández, I., Hill, F., Bogart, R., Rabello-Soares, M. C. & Haber, D. 2013 Subsurface Meridional Flow from HMI Using the Ring-Diagram Pipeline. *Sol. Phys.* **287**, 85–106.

Komm, R. W., Howard, R. F. & Harvey, J. W. 1993 Meridional Flow of Small Photospheric Magnetic Features. *Sol. Phys.* **147**, 207–223.

Komm, R. W., Howard, R. F. & Harvey, J. W. 1995 Characteristic size and diffusion of quiet Sun magnetic patterns. *Sol. Phys.* **158**, 213–225.

Krause, F. & Rädler, K.-H. 1980 *Mean-field magnetohydrodynamics and dynamo theory*. Pergamon Press.

Kumar, R., Jouve, L. & Nandy, D. 2019 A 3D kinematic Babcock Leighton solar dynamo model sustained by dynamic magnetic buoyancy and flux transport processes. *A&A* **623**, A54.

Kumar, R., Jouve, L., Pinto, R. F. & Rouillard, A. P. 2018 Production of sunspots and their effects on the corona and solar wind: Insights from a new 3D flux-transport dynamo model. *Frontiers in Astronomy and Space Sciences* **5**, 4.

Larmor, J. 1919 How could a rotating body such as the Sun become a magnet? *Report of the British Association for the Advancement of Science* **87th Meeting**, 159–160.

Layden, A. C., Fox, P. A., Howard, J. M., Sarajedini, A., Schatten, K. H. & Sofia, S. 1991 Dynamo-based scheme for forecasting the magnitude of solar activity cycles. *Sol. Phys.* **132**, 1–40.

Leighton, R. 1965 Introductory report. In *Stellar and Solar Magnetic Fields* (ed. R. Lust), *IAU Symposium*, vol. 22, p. 158.

Leighton, R. B. 1964 Transport of Magnetic Fields on the Sun. *ApJ* **140**, 1547.

Lemerle, A. & Charbonneau, P. 2017 A Coupled 2 x 2D Babcock-Leighton Solar Dynamo Model. II. Reference Dynamo Solutions. *ApJ* **834**, 133.

Lemerle, A., Charbonneau, P. & Carignan-Dugas, A. 2015 A Coupled 2 x 2D Babcock-Leighton Solar Dynamo Model. I. Surface Magnetic Flux Evolution. *ApJ* **810**, 78.

Li, J. & Ulrich, R. K. 2012 Long-term Measurements of Sunspot Magnetic Tilt Angles. *ApJ* **758**, 115.

Li, K. J., Yun, H. S. & Gu, X. M. 2001 On long-term predictions of the maximum sunspot numbers of solar cycles 21 to 23. *A&A* **368**, 285–291.

Mackay, D. H., Jardine, M., Collier Cameron, A., Donati, J. F. & Hussain, G. A. J. 2004 Polar caps on active stars: magnetic flux emergence and transport. *MNRAS* **354** (3), 737–752.

Mackay, D. H. & Yeates, A. R. 2012 The Sun's Global Photospheric and Coronal Magnetic Fields: Observations and Models. *LRSP* **9**, 6.

Mamajek, E. E., Prsa, A., Torres, G., Harmanec, P., Asplund, M., Bennett, P. D., Capitaine, N., Christensen-Dalsgaard, J., Depagne,

E., Folkner, W. M., Haberreiter, M., Hekker, S., Hilton, J. L., Kostov, V., Kurtz, D. W., Laskar, J., Mason, B. D., Milone, E. F., Montgomery, M. M., Richards, M. T., Schou, J. & Stewart, S. G. 2015 IAU 2015 Resolution B3 on Recommended Nominal Conversion Constants for Selected Solar and Planetary Properties. *ArXiv e-prints* .

Mardia, K., Kent, J. & Bibby, J. 1979 *Multivariate Analysis*. Acad. Press, London [u.a.].

Martin-Belda, D. & Cameron, R. H. 2016 Surface flux transport simulations: Effect of inflows toward active regions and random velocities on the evolution of the Sun's large-scale magnetic field. *A&A* **586**, A73.

Martínez Pillet, V., Lites, B. W. & Skumanich, A. 1997 Active Region Magnetic Fields. I. Plage Fields. *ApJ* **474**, 810–842.

McIntosh, P. S., Brown, G. M., Buhmann, R., Clark, T., Fougere, P. F., Hunter, H., Lincoln, J. V., Sargent, III, H. H., Timothy, J. G. & Lin, Y. Z. 1979 Long-term solar activity predictions. In *NOAA Solar-Terrestrial Predictions Proceedings. Volume 2.* (ed. R. F. Donnelly), *Solar-Terrestrial Physics*, vol. 2.

Metcalfe, T. S. & Charbonneau, P. 2003 Stellar structure modeling using a parallel genetic algorithm for objective global optimization. *Journal of Computational Physics* **185**, 176–193.

Miesch, M. S. & Dikpati, M. 2014 A Three-dimensional Babcock-Leighton Solar Dynamo Model. *ApJL* **785**, L8.

Miesch, M. S. & Teweldebirhan, K. 2016 A three-dimensional Babcock-Leighton solar dynamo model: Initial results with axisymmetric flows. *Advances in Space Research* **58**, 1571–1588.

Moffatt, H. K. 1978 *Magnetic field generation in electrically conducting fluids*. Cambridge University Press.

Mosher, J. M. 1977 *The magnetic history of solar active regions*. PhD Thesis, California Institute of Technology, Pasadena.

Muñoz-Jaramillo, A., Dasi-Espuig, M., Balmaceda, L. A. & DeLuca, E. E. 2013 Solar Cycle Propagation, Memory, and Prediction: Insights from a Century of Magnetic Proxies. *ApJL* **767**, L25.

Muñoz-Jaramillo, A., Nandy, D. & Martens, P. C. 2008 Effect of the Magnetic Quenching of the Turbulent Diffusivity in a Mean-Field Kinematic Solar Dynamo. *AGU Spring Meeting Abstracts* pp. SP41A–09.

Muñoz-Jaramillo, A., Nandy, D. & Martens, P. C. H. 2009 Helioseismic Data Inclusion in Solar Dynamo Models. *ApJ* **698**, 461–478.

Muñoz-Jaramillo, A., Nandy, D. & Martens, P. C. H. 2011 Magnetic Quenching of Turbulent Diffusivity: Reconciling Mixing-length Theory Estimates with Kinematic Dynamo Models of the Solar Cycle. *ApJL* **727**, L23.

Muñoz-Jaramillo, A., Nandy, D., Martens, P. C. H. & Yeates, A. R. 2010 A Double-ring Algorithm for Modeling Solar Active Regions: Unifying Kinematic Dynamo Models and Surface Flux-transport Simulations. *ApJL* **720**, L20–L25.

Muñoz-Jaramillo, A., Senkpeil, R. R., Longcope, D. W., Tlatov, A. G., Pevtsov, A. A., Balmaceda, L. A., DeLuca, E. E. & Martens, P. C. H. 2015 The Minimum of Solar Cycle 23: As Deep as It Could Be? *ApJ* **804**, 68.

Muñoz-Jaramillo, A., Sheeley, N. R., Zhang, J. & DeLuca, E. E. 2012 Calibrating 100 Years of Polar Faculae Measurements: Implications for the Evolution of the Heliospheric Magnetic Field. *ApJ* **753**, 146.

Nagy, M., Lemerle, A., Labonville, F., Petrovay, K. & Charbonneau, P. 2017 The Effect of "Rogue" Active Regions on the Solar Cycle. *Sol. Phys.* **292**, 167.

Nandy, D. & Choudhuri, A. R. 2001 Toward a Mean Field Formulation of the Babcock-Leighton Type Solar Dynamo. I. α-Coefficient versus Durney's Double-Ring Approach. *ApJ* **551**, 576–585.

Nandy, D. & Choudhuri, A. R. 2002 Explaining the Latitudinal Distribution of Sunspots with Deep Meridional Flow. *Science* **296** (5573), 1671–1673.

Nita, G., Angryk, R., Aydin, B., Banda, J., Bastian, T., Berger, T., Bindi, V., Boucheron, L., Cao, W., Christian, E., de Nolfo, G., DeLuca, E., DeRosa, M., Downs, C., Fleishman, G., Fuentes, O., Gary, D., Hill, F., Hoeksema, T., Hu, Q., Ilie, R., Ireland, J., Kamalabadi, F., Korreck, K., Kosovichev, A., Lin, J., Lugaz, N., Mannucci, A., Mansour, N., Martens, P., Mays, L., McAteer, J., McIntosh, S. W., Oria, V., Pan, D., Panesi, M., Pesnell, W. D., Pevtsov, A., Pillet, V., Rachmeler, L., Ridley, A., Scherliess, L., Toth, G., Velli, M., White, S., Zhang, J. & Zou, S. 2018 Roadmap for Reliable Ensemble Forecasting of the Sun-Earth System. *arXiv e-prints* .

Parker, E. N. 1955*a* Hydromagnetic Dynamo Models. *ApJ* **122**, 293.

Parker, E. N. 1955*b* The Formation of Sunspots from the Solar Toroidal Field. *ApJ* **121**, 491.

Parker, E. N. 1979 *Cosmical magnetic fields: Their origin and their activity*. Clarendon Press.

Pesnell, W. D. 2016 Predictions of solar cycle 24: How are we doing? *Space Weather* **14** (1), 10–21.

Petrie, G. J. D. 2012 Evolution of Active and Polar Photospheric Magnetic Fields During the Rise of Cycle 24 Compared to Previous Cycles. *Sol. Phys.* **281**, 577–598.

Pevtsov, A. A., Virtanen, I., Mursula, K., Tlatov, A. & Bertello, L. 2016 Reconstructing solar magnetic fields from historical observations. I. Renormalized Ca K spectroheliograms and pseudo-magnetograms. *A&A* **585**, A40.

Prandtl, L. 1925 Bericht uber untersuchungen zur ausgebildeten turbulenz. *Zs. angew. Math. Mech.* **5**, 136–139.

Priest, E. R. 1982 *Solar magneto-hydrodynamics*. D. Reidel Publishing Company.

Pukelsheim, F. 1994 The Three Sigma Rule. *The American Statistician* **48**, 88–91.

Rast, M. P. 2003 The Scales of Granulation, Mesogranulation, and Supergranulation. *ApJ* **597**, 1200–1210.

Rempel, M. 2007 Origin of Solar Torsional Oscillations. *ApJ* **655**, 651–659.

Rieutord, M. & Rincon, F. 2010 The Sun's Supergranulation. *LRSP* **7**, 2.

Riley, P., Ben-Nun, M., Linker, J. A., Mikic, Z., Svalgaard, L., Harvey, J., Bertello, L., Hoeksema, T., Liu, Y. & Ulrich, R. 2014 A Multi-Observatory Inter-Comparison of Line-of-Sight Synoptic Solar Magnetograms. *Sol. Phys.* **289**, 769–792.

Roberts, P. H. & Soward, A. M. 1975 A unified approach to mean field electrodynamics. *Astronomische Nachrichten* **296**, 49–64.

Sacks, J., Welch, W. J., Mitchell, T. J. & Wynn, H. P. 1989 Design and Analysis of Computer Experiments. *Statistical Science* **4** (4), 409–423.

Santner, T. J., Williams, B. J. & Notz, W. I. 2003 *The Design and Analysis of Computer Experiments*. Springer-Verlag New York.

Schatten, K. H., Scherrer, P. H., Svalgaard, L. & Wilcox, J. M. 1978 Using dynamo theory to predict the sunspot number during solar cycle 21. *Geophys. Res. Lett.* **5**, 411–414.

Scherrer, P. H., Bogart, R. S., Bush, R. I., Hoeksema, J. T., Kosovichev, A. G., Schou, J., Rosenberg, W., Springer, L., Tarbell, T. D., Title, A., Wolfson, C. J., Zayer, I. & MDI Engineering Team 1995 The Solar Oscillations Investigation - Michelson Doppler Imager. *Sol. Phys.* **162**, 129–188.

Schou, J., Antia, H. M., Basu, S., Bogart, R. S., Bush, R. I., Chitre, S. M., Christensen-Dalsgaard, J., Di Mauro, M. P., Dziembowski, W. A., Eff-Darwich, A., Gough, D. O., Haber, D. A., Hoeksema, J. T., Howe, R., Korzennik, S. G., Kosovichev, A. G., Larsen, R. M., Pijpers, F. P., Scherrer, P. H., Sekii, T., Tarbell, T. D., Title, A. M., Thompson, M. J. & Toomre, J. 1998 Helioseismic Studies of Differential Rotation in the Solar Envelope by the Solar Oscillations Investigation Using the Michelson Doppler Imager. *ApJ* **505**, 390–417.

Schrijver, C. J. 2001 Simulations of the Photospheric Magnetic Activity and Outer Atmospheric Radiative Losses of Cool Stars Based on Characteristics of the Solar Magnetic Field. *ApJ* **547**, 475–490.

Schrijver, C. J. 2015 Socio-Economic Hazards and Impacts of Space Weather: The Important Range Between Mild and Extreme. *Space Weather* **13**, 524–528.

Schrijver, C. J. & Beer, J. 2014 Space Weather From Explosions on the Sun: How Bad Could It Be? *EOS Transactions* **95**, 201–202.

Schrijver, C. J., De Rosa, M. L. & Title, A. M. 2002 What Is Missing from Our Understanding of Long-Term Solar and Heliospheric Activity? *ApJ* **577**, 1006–1012.

Schrijver, C. J., Kauristie, K., Aylward, A. D., Denardini, C. M., Gibson, S. E., Glover, A., Gopalswamy, N., Grande, M., Hapgood, M., Heynderickx, D., Jakowski, N., Kalegaev, V. V., Lapenta, G., Linker, J. A., Liu, S., Mandrini, C. H., Mann, I. R., Nagatsuma, T., Nandy, D., Obara, T., Paul O'Brien, T., Onsager, T., Opgenoorth, H. J., Terkildsen, M., Valladares, C. E. & Vilmer, N. 2015 Understanding space weather to shield society: A global road map for 2015-2025 commissioned by COSPAR and ILWS. *Advances in Space Research* **55**, 2745–2807.

Schrijver, C. J. & Liu, Y. 2008 The Global Solar Magnetic Field Through a Full Sunspot Cycle: Observations and Model Results. *Sol. Phys.* **252**, 19–31.

Schrijver, C. J. & Title, A. M. 2001 On the Formation of Polar Spots in Sun-like Stars. *ApJ* **551**, 1099–1106.

Schüssler, M. & Baumann, I. 2006 Modeling the Sun's open magnetic flux. *A&A* **459**, 945–953.

Schüssler, M. & Rempel, M. 2005 The dynamical disconnection of sunspots from their magnetic roots. *A&A* **441**, 337–346.

Schwabe, M. 1844 Sonnenbeobachtungen im Jahre 1843. Von Herrn Hofrath Schwabe in Dessau. *Astronomische Nachrichten* **21**, 233.

Sheeley, N. R., J., Boris, J. P., Young, T. R., J., DeVore, C. R. & Harvey, K. L. 1983 A Quantitative Study of Magnetic Flux Transport on the Sun. In *Solar and Stellar Magnetic Fields: Origins and Coronal Effects* (ed. J. O. Stenflo), *IAU Symposium*, vol. 102, p. 273.

Sheeley, Jr., N. R. 2005 Surface Evolution of the Sun's Magnetic Field: A Historical Review of the Flux-Transport Mechanism. *LRSP* **2**, 5.

Snodgrass, H. B. & Ulrich, R. K. 1990 Rotation of Doppler features in the solar photosphere. *ApJ* **351**, 309–316.

Spiegel, E. A. & Zahn, J.-P. 1992 The solar tachocline. *A&A* **265**, 106–114.

Spörer, G. 1880 Beobachtungen der Sonnenflecken von Januar 1874 bis December 1879. *Publikationen des Astrophysikalischen Observatoriums zu Potsdam* **2**, 1–81.

Svalgaard, L., Duvall, Jr., T. L. & Scherrer, P. H. 1978 The strength of the sun's polar fields. *Sol. Phys.* **58**, 225–239.

Thibault, K., Charbonneau, P. & Béland, M. 2014 Solar Photospheric Network Properties and Their Cycle Variation. *ApJ* **796**, 19.

Tobias, S. M. 1996 Diffusivity Quenching as a Mechanism for Parker's Surface Dynamo. *ApJ* **467**, 870.

Tobias, S. M. 2002 The solar dynamo. *Philosophical Transactions of the Royal Society of London Series A* **360**, 2741–2756.

Tobias, S. M., Brummell, N. H., Clune, T. L. & Toomre, J. 1998 Pumping of Magnetic Fields by Turbulent Penetrative Convection. *ApJL* **502**, L177–L180.

Topka, K., Moore, R., Labonte, B. J. & Howard, R. 1982 Evidence for a poleward meridional flow on the sun. *Sol. Phys.* **79**, 231–245.

Ulrich, R. K. 2010 Solar Meridional Circulation from Doppler Shifts of the Fe I Line at 5250 Å as Measured by the 150-foot Solar Tower Telescope at the Mt. Wilson Observatory. *ApJ* **725**, 658–669.

Upton, L. & Hathaway, D. H. 2014*a* Effects of Meridional Flow Variations on Solar Cycles 23 and 24. *ApJ* **792**, 142.

Upton, L. & Hathaway, D. H. 2014*b* Predicting the Sun's Polar Magnetic Fields with a Surface Flux Transport Model. *ApJ* **780**, 5.

Upton, L. A. & Hathaway, D. H. 2018 An Updated Solar Cycle 25 Prediction with AFT: The Modern Minimum. *ArXiv e-prints* .

van Ballegooijen, A. A. 1998 Understanding the Solar Cycle. In *Synoptic Solar Physics* (ed. K. S. Balasubramaniam, J. Harvey & D. Rabin), *Astronomical Society of the Pacific Conference Series*, vol. 140, p. 17.

van Ballegooijen, A. A., Cartledge, N. P. & Priest, E. R. 1998 Magnetic Flux Transport and the Formation of Filament Channels on the Sun. *ApJ* **501**, 866–881.

van Ballegooijen, A. A. & Mackay, D. H. 2007 Model for the Coupled Evolution of Subsurface and Coronal Magnetic Fields in Solar Active Regions. *ApJ* **659**, 1713–1725.

van Ballegooijen, A. A., Priest, E. R. & Mackay, D. H. 2000 Mean Field Model for the Formation of Filament Channels on the Sun. *ApJ* **539**, 983–994.

van Driel-Gesztelyi, L. & Culhane, J. L. 2009 Magnetic Flux Emergence, Activity, Eruptions and Magnetic Clouds: Following Magnetic Field from the Sun to the Heliosphere. *Space Sci. Rev.* **144**, 351–381.

Vaquero, J. M. 2007 Letter to the Editor: Sunspot observations by Theophrastus revisited. *Journal of the British Astronomical Association* **117**, 346.

Vennerstrøm, S. & Friis-Christensen, E. 1987 On the role of IMF By in generating the electric field responsible for the flow across the polar cap. *J. Geophys. Res.* **92**, 195–202.

Vernon, I., Goldstein, M. & Bower, R. 2010 Galaxy Formation: a Bayesian Uncertainty Analysis. *Bayesian Analysis* **5** (4), 619–670.

Vernon, I., Goldstein, M. & Bower, R. 2014 Galaxy Formation: Bayesian History Matching for the Observable Universe. *Statistical Science* **29** (1), 81–90.

Vernon, I., Liu, J., Goldstein, M., Rowe, J., Topping, J. & Lindsey, K. 2018 Bayesian uncertainty analysis for complex systems biology models: emulation,

global parameter searches and evaluation of gene functions. *BMC Systems Biology* **12** (1), 1–29.

Virtanen, I. O. I., Virtanen, I. I., Pevtsov, A. A. & Mursula, K. 2018 Reconstructing solar magnetic fields from historical observations. III. Activity in one hemisphere is sufficient to cause polar field reversals in both hemispheres. *A&A* **616**, A134.

Virtanen, I. O. I., Virtanen, I. I., Pevtsov, A. A., Yeates, A. & Mursula, K. 2017 Reconstructing solar magnetic fields from historical observations. II. Testing the surface flux transport model. *A&A* **604**, A8.

Wang, Y.-M. 2017 Surface Flux Transport and the Evolution of the Sun's Polar Fields. *Space Sci. Rev.* **210**, 351–365.

Wang, Y.-M., Lean, J. & Sheeley, Jr., N. R. 2002*a* Role of a Variable Meridional Flow in the Secular Evolution of the Sun's Polar Fields and Open Flux. *ApJL* **577**, L53–L57.

Wang, Y.-M., Nash, A. G. & Sheeley, Jr., N. R. 1989*a* Evolution of the sun's polar fields during sunspot cycle 21 - Poleward surges and long-term behavior. *ApJ* **347**, 529–539.

Wang, Y.-M., Nash, A. G. & Sheeley, Jr., N. R. 1989*b* Magnetic flux transport on the sun. *Science* **245**, 712–718.

Wang, Y. M., Sheeley, N. R., J., Nash, A. G. & Shampine, L. R. 1988 The Quasi-rigid Rotation of Coronal Magnetic Fields. *ApJ* **327**, 427.

Wang, Y.-M. & Sheeley, Jr., N. R. 1989 Average properties of bipolar magnetic regions during sunspot cycle 21. *Sol. Phys.* **124**, 81–100.

Wang, Y.-M. & Sheeley, Jr., N. R. 1991 Magnetic flux transport and the sun's dipole moment - New twists to the Babcock-Leighton model. *ApJ* **375**, 761–770.

Wang, Y.-M., Sheeley, Jr., N. R. & Lean, J. 2002*b* Meridional Flow and the Solar Cycle Variation of the Sun's Open Magnetic Flux. *ApJ* **580**, 1188–1196.

Wang, Y.-M., Sheeley, Jr., N. R. & Nash, A. G. 1991 A new solar cycle model including meridional circulation. *ApJ* **383**, 431–442.

Whitbread, T. 2019 Active region dipole moments determined from NSO carrington maps.

Whitbread, T., Yeates, A. R. & Muñoz-Jaramillo, A. 2018 How Many Active Regions Are Necessary to Predict the Solar Dipole Moment? *ApJ* **863**, 116.

Whitbread, T., Yeates, A. R. & Muñoz-Jaramillo, A. 2019 The need for active region disconnection in 3D kinematic dynamo simulations. *A&A* **627**, A168.

Whitbread, T., Yeates, A. R., Muñoz-Jaramillo, A. & Petrie, G. J. D. 2017 Parameter optimization for surface flux transport models. *A&A* **607**, A76.

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L. & Yamazaki, K. 2013 History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics* **41** (7), 1703–1729.

Wilmot-Smith, A. L., Priest, E. R. & Hornig, G. 2005 Magnetic diffusion and the motion of field lines. *Geophysical and Astrophysical Fluid Dynamics* **99**, 177–197.

Wolf, R. & Brunner, W. 1935 Neue Eigenschaften der Sonnenfleckenkurve. *Astronomische Mitteilungen der Eidgenössischen Sternwarte Zurich* **14**, 105–136.

Worden, J. & Harvey, J. 2000 An Evolving Synoptic Magnetic Flux map and Implications for the Distribution of Photospheric Magnetic Flux. *Sol. Phys.* **195**, 247–268.

Yeates, A. R. 2014 Coronal Magnetic Field Evolution from 1996 to 2012: Continuous Non-potential Simulations. *Sol. Phys.* **289**, 631–648.

Yeates, A. R. 2016 Bipolar magnetic regions determined from NSO synoptic carrington maps.

Yeates, A. R., Baker, D. & van Driel-Gesztelyi, L. 2015 Source of a Prominent Poleward Surge During Solar Cycle 24. *Sol. Phys.* **290**, 3189–3201.

Yeates, A. R. & Mackay, D. H. 2012 Chirality of High-latitude Filaments over Solar Cycle 23. *ApJL* **753**, L34.

Yeates, A. R., Mackay, D. H. & van Ballegooijen, A. A. 2007 Modelling the Global Solar Corona: Filament Chirality Observations and Surface Simulations. *Sol. Phys.* **245**, 87–107.

Yeates, A. R., Mackay, D. H., van Ballegooijen, A. A. & Constable, J. A. 2010 A nonpotential model for the Sun's open magnetic flux. *J. Geophys. Res. (Space Phys.)* **115**, A09112.

Yeates, A. R. & Muñoz-Jaramillo, A. 2013 Kinematic active region formation in a three-dimensional solar dynamo model. *MNRAS* **436**, 3366–3379.

Zhao, J. & Kosovichev, A. G. 2004 Torsional Oscillation, Meridional Flows, and Vorticity Inferred in the Upper Convection Zone of the Sun by Time-Distance Helioseismology. *ApJ* **603**, 776–784.